

Computer leert wijktypen bepalen voor heel Nederland

een studie naar het inzetten van kunstmatige intelligentie voor het bepalen van wijktypen

Author(s)

Kluck, Jeroen; Kleerekoper, L.aura; Erwin, Stephanie; Coppel, Lisanne; Bons, Pieter; Arif, Zerin; Geisler, Len; Veenbos, Koen; Koekoek, Arjen

Publication date

2023

Document Version

Final published version

License

CC BY-NC-SA

[Link to publication](#)

Citation for published version (APA):

Kluck, J., Kleerekoper, L. A., Erwin, S., Coppel, L., Bons, P., Arif, Z., Geisler, L., Veenbos, K., & Koekoek, A. (2023). *Computer leert wijktypen bepalen voor heel Nederland: een studie naar het inzetten van kunstmatige intelligentie voor het bepalen van wijktypen*. Climate Adaptation Services.



General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: <https://www.amsterdamuas.com/library/contact>, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Computer leert wijktypen bepalen voor heel Nederland

Een studie naar het inzetten van kunstmatige intelligentie voor het bepalen van wijktypen



Auteurs Hogeschool van Amsterdam:

Jeroen Kluck
Laura Kleerekoper
Stephanie Erwin
Lisanne Corpel
Pieter Bons
Zerin Arif

Tauw B.V.: Len Geisler

Opdrachtgever: Stichting CAS
Koen Veenenbos
Arjen Koekoek

Datum: September 2023

Inhoud

1. Inleiding.....	4
1.1. Aanleiding	5
1.2. Doelstelling.....	7
1.3. Aanpak	7
1.4. Leeswijzer.....	7
2. Methode bepalen wijktypen met behulp van kunstmatige intelligentie	8
2.1. Methodiek training en toetsing aan de hand van expertvalidatie.....	8
2.2. Data testen voor vijf testgemeenten	9
3. Data voor wijktypen	9
3.1. Datasets overzicht.....	9
3.2. Geoprocessing datasets	14
3.2.1. Bebouwde kom.....	14
3.2.2. Bouwjaar	14
3.2.3. Bouwhoogte	15
3.2.4. Woningtype	15
3.2.5. Daktype.....	15
3.2.6. Netwerkanalyse	15
3.2.7. Huishoudens per hectare.....	17
3.2.8. Dichtheid/Stedelijkheid	18
3.2.9. Verharding	18
3.2.10 Wonen	18
3.2.11. Groen	19
3.2.12. Rudifun	19
4. Machine learning inzetten voor het bepalen van wijktypen	19
4.1. Machine Learning modellering met het Random Forest Model.....	19
4.2 Relevantie van de datasets	21
4.3. Confusion Matrix voor de wijktipe bepaling	21
4.3. Nauwkeurigheid bepaling wijktypen door het model.....	22
5. Clustering van PC6 gebieden	25
5.1. Beslisregels wijktypen toekennen per buurt.....	26
6. Conclusies.....	27
7. Discussie	29
7.1. Beperkingen van methodiek	29
7.2. Aanbevelingen voor verbeteringen.....	30
Literatuurlijst.....	32
Bijlage 1.....	33
Bijlage 2.....	34



1. Inleiding

Het beschrijven van een bebouwd gebied kan in algemene termen, zoals 'in het centrum' of 'in een groene wijk'. Voor een meer specifieke beschrijving geeft alleen de aanduiding van een wijktype, zoals historische binnenstad, volkswijk of naoorlogse tuinstad veel informatie over het bebouwd gebied. Met de aanduiding van een wijktype worden kenmerken direct duidelijk. Denk aan bouwperiode, bouwstijl, mate van verstedelijking, type en grootte van woningen, soort en hoeveelheid groen en water, de indeling van de openbare ruimte en het wegpatroon. De specifieke combinatie van deze ruimtelijke eigenschappen beïnvloeden de kwetsbaarheid voor klimaatverandering en de mogelijkheden voor klimaatadaptatie. Dit kan ontwerpers en planners helpen keuzes te maken voor een specifieke locatie op basis van het wijktype.

Er is behoefte aan kennis over de kwetsbaarheid van stadsbewoners door klimaatverandering en gerichte oplossingen voor een brede groep professionals. Het beroepenveld van beleidsmakers, stedenbouwkundigen en bijvoorbeeld planologen, dat zich richt op de openbare ruimte, heeft niet altijd de specialistische kennis over watermanagement of het stedelijk microklimaat die nodig is om keuzes voor klimaatadaptatiemaatregelen te maken. Inrichting van de openbare ruimte is vaak een complexe opgave met vele betrokken partijen met uiteenlopende achtergronden. Door klimaatadaptatiemaatregelen vroeg in het proces mee te nemen is de kans op mee-koppeling groter en daarmee ook de kans op uitvoering en draagvlak. Om klimaatadaptatie vroeg in het ontwerpproces mee te kunnen nemen biedt de indeling van het bebouwd gebied in wijktypen en daarbij passende maatregelen de mogelijkheid een voorselectie te maken.

Met de indeling van bebouwd gebied in wijktypen kunnen aan de ene kant algemene kwetsbaarheden voor ontwikkelingen die niet klimaat robuust zijn inzichtelijk worden gemaakt. Zo blijken tuindorpen vooral kwetsbaar voor verstening van privé tuinen omdat het groen in deze wijken voor het overgrote deel particulier is. En een kwetsbaarheid van bloemkoolwijken is bijvoorbeeld de ruimte die de auto zich in deze wijken heeft toegeëigend waardoor straten en binnenhoven veel verharding hebben.

Aan de andere kant maken de wijktypen het ook mogelijk generieke maatregelen voor klimaatadaptatie binnen dezelfde typologie op te stellen en te presenteren. De kenmerken van een wijktype bepalen hoe er in de straat rekening kan worden gehouden met een extremer klimaat. Zo biedt het vele publieke groen in naoorlogse tuinsteden ruimte om water bovengronds te bergen om problemen met extreme neerslag te voorkomen. De stedelijke bouwblokken daarentegen vragen eerder om (technische) oplossingen ondergronds. En veel bloemkoolwijken uit de jaren '70 zijn rijk aan groen rondom de huizen dat met een verbeterde toegankelijkheid een koele plek zou kunnen bieden tijdens warme dagen.

Het vakgebied dat zich richt op inrichting van de openbare ruimte gebruikt in de communicatie vaak een indeling in wijktypen. Deze groep kan de wijktypenkaart inzetten om analyses en onderzoek op grotere schaal te doen naar kwetsbaarheden en oplossingen. Dit is bijvoorbeeld gedaan in het hittebestendige stad onderzoek voor de tien wijktypen in [Figuur 1](#).

Iedereen die minder direct werkt met wijktypen in de praktijk, kan eenvoudig met de kaart de wijktypen onderscheiden en gebruik maken van de algemene kennis die over een wijktype is gegenereerd.



Figuur 1: Indeling in wijktypen in De hittebestendige stad: Coolkit, een toolkit voor ontwerpers van de buitenruimte (Kluck, et al. 2020)

1.1. Aanleiding

Voor de Klimaat-effectatlas van stichting Climate Adaptation Services (CAS) is in 2017 een onderverdeling van wijktypen gemaakt die aansluit bij klimaatbestendig inrichten welke zijn gebaseerd op de classificering in het proefschrift van Kleerekoper (2016). De indeling van de wijktypen wordt in veel analyses en presentaties over klimaatadaptatie gebruikt. De kaart uit 2017 diende te worden verbeterd omdat:

1. De kaart enkel op buurtniveau beschikbaar was, waardoor detailniveau ontbrak. Zo werd de variatie die in een dorp of buurt aanwezig is niet goed weergegeven.
2. Een aantal wijktypen werden niet goed bepaald in de handmatige GIS-methode. Met name veel wijken waren gedefinieerd als bloemkoolwijk. De onderscheiding tussen 'Bloemkoolwijk' en 'Naoorlogse woonwijk' was niet goed genoeg.

Een toevoeging aan de nieuwe wijktipekaart is een indicatie van de kwetsbaarheid voor hittestress per wijktipe, zoals in de voorloper voor wateroverlast is gedaan. Zie ook [Figuur 2](#).

	DOMINANTE WIJKTYPOLOGIE	BOUWPERIODE	KENMERKEN	INDICATIEVE KWETSBAARHEID WATEROVERLAST
	Historische binnenstad	voor 1900	veel verharding, 3-5 lagen, monumentaal groen	☹☹☹
	Stedelijk bouwblok	voor 1940	geen voortuin of groenstrook, 4-8 lagen	☹☹☹
	Vooroorlogs bouwblok	1900-1940	niet altijd voortuin, 3-4 lagen, bredere straten dan stedelijk bouwblok en soms groenstrook	☹☹☹
	Tuindorp	1910-1940	ruime voor- en achtertuinen, 2-3 lagen, veel langsparkeren, jaren 30-bouwstijl, beperkt gemeentelijk groen, vaak geen straatbomen	☹☹☹
	Volkswijk	1910-1940	geen voortuin, weinig gemeentelijk groen, 2-3 lagen, eengezinswoningen	☹☹
	Tuinstad laagbouw	1945-1960	open bouwblokken met veel groen, 2-3 lagen, eengezinswoningen	☹☹
	Tuinstad hoogbouw	1945-1970	open bouwblokken met veel groen, 4-6 lagen, appartementen, berging op begane grond	☹☹
	Naoorlogse woonwijk	1945-1990	voor- en achtertuin, 2-3 lagen, eengezinswoningen in rij, zonder1 kap of vrijstaand	☹☹
	Bloemkoolwijk	1970-1990	eengezinswoningen met voor- en achtertuin, kronkelende stratenpatronen, hofjes, brede groenstrook rondom de wijk	☹
	Hoogbouw	1945-heden	meer dan 10 lagen, gebouwen in grid	☹☹
	Sub-urbane uitbreiding – Vinex	1990-heden	eengezinswoningen in rij, twee onder een kap, vrijstaand, appartementen	☹
	Vernieuwd	1990-heden	Vernieuwing bestaande bouw, veelal hoge dichtheden	☹
	Villa	Van alle tijden	Veel ruimte tussen woningen, losstaande huizen	☹
	Bedrijven	Van alle tijden	Bedrijvigheid	☹

Figuur 2: Legenda van de wijktipekaart zoals deze in 2017 is verschenen op www.klimaat-effectatlas.nl

De kaart is nu verbeterd door gebruik te maken van nieuwe databronnen en computertechnieken. Dit heeft geleid tot een innovatieve methodiek voor het bepalen van de volgende wijktypologieën: Historische binnenstad, stedelijk bouwblok, vooroorlogse woonwijk, tuindorp, volkswijk, tuinstad laagbouw, tuinstad hoogbouw, naoorlogse woonwijk, bloemkoolwijk, hoogbouw, sub-urbane uitbreiding/Vinex, vernieuwd, villa, vooroorlogse woonwijk, bedrijven.

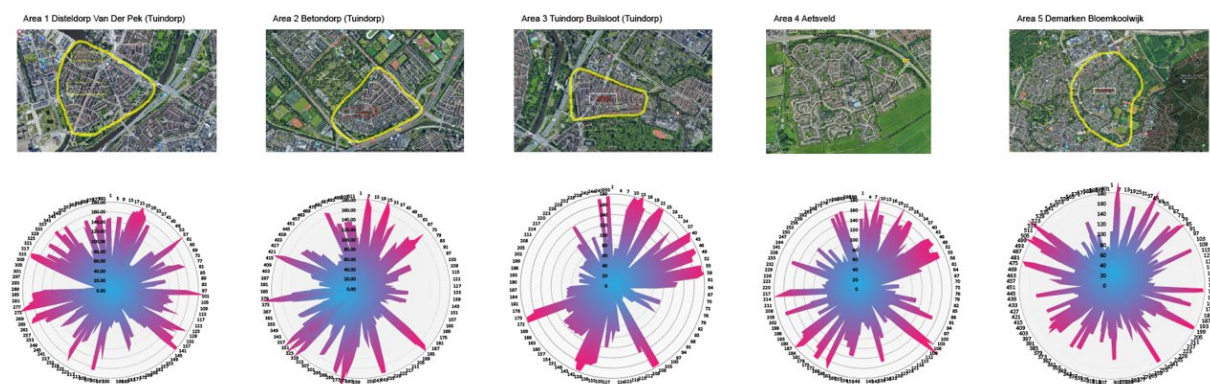
Potentie voor een nieuwe kaart

In de aanloop naar de nieuwe wijktypenkaart zijn een aantal vooronderzoeken gedaan. Er is een verkenning gedaan door studenten van de minor Urban Analytics en afstudeerder Wiskunde. Parallel aan het werk van de studenten werkten onderzoekers van het Lectoraat Water in en om de stad aan dezelfde vragen en hebben meerdere methoden verkend en uitgetest. De uitkomsten toonden aan dat er voldoende potentie was voor een verbeterslag. De twee belangrijkste verbeterpunten hierin waren:

1. Nieuwe clustering op basis van zes-posities postcodevlakken (PC6 gebieden, zoals 1234AB), dit zorgt voor logische keuze in grenzen voor clusters kleiner dan de buurtgrens.
2. Nieuwe methoden en data om specifieke wijktypen van elkaar te onderscheiden.

De eerste verbeterslag betreft een verbeterde clustering van PC6 gebieden. Hiervoor zijn verschillende methodes getest met elk een gebrek. Zoals het clusteringsmodel 'Nearest Neighbour' waarbij automatisch groepen die van hetzelfde type zijn en aan elkaar grenzen samen worden gebundeld. Voor homogene gebieden werkt dat aardig, maar bij gemengde wijken geeft het een versplinterd beeld. De conclusie is dat dit clusteringsmodel de huidige buurtkaart niet voldoende verbeterd. Er zijn manieren om automatisch bij gemengde buurten en grenzen tussen clusters losse PC6 gebieden op te nemen in een groter cluster. Daarbij ontstaan situaties waarin de clusters onlogische uitkomsten geven voor met name de randen van de clusters. Daarom is ervoor gekozen om deze manier van clusteren niet verder te onderzoeken en te kijken naar de bestaande grenzen van PC6, 5 en 4. Waarbij PC5 het meest toevoegt naast de PC6 en buurtniveaus.

De tweede verbeterslag betreft het toevoegen van data die helpen in het onderscheiden van de wijktypen. Om bijvoorbeeld de Bloemkoolwijken beter te onderscheiden is vooral gekeken naar netwerkstructuren. In het wegenpatroon zijn de volgende variabelen onderzocht, zoals de oriëntatie van wegen (Figuur 3), straatlengte en aantal wegen per kruispunt. Het blijkt dat alle eigenschappen samen tot een nauwkeuriger beslissing leiden. Voor het onderscheid tussen tuindorpen en volkswijken is gekeken naar schuine daken en kleinere stapjes in gebouwhoogte in plaats van het aantal bouwlagen. Het is nog niet gelukt om een indicator te vinden die voortuinen kan identificeren. Maar ook zonder deze toevoeging wordt de inschatting voor tuindorpen al sterk verbeterd. In de voorverkenning blijken ook extra variabelen zoals mate van stedelijkheid, FSI (Floor Space Index) GSI (Ground Space Index) en OSR (Open Space Ratio) voor verbetering te zorgen.



Figuur 3: Oriëntatie van straten in een wijk ter voorverkenning van deze variabele.

1.2. Doelstelling

De doelstelling voor de nieuwe wijktypekaart is tweeledig: met meer nauwkeurigheid het wijktype op PC6 niveau kunnen genereren voor heel Nederland en daarnaast ook een opdeling van buurten te kunnen geven zodat gemengde wijken met een logische indeling in sub-buurten worden weergegeven.

De opdeling in wijktypen dient in eerste instantie ter ondersteuning in het testen en kiezen van maatregelen voor klimaatadaptatie in een veralgemeniseerd wijktype, maar kunnen ook worden ingezet in vele andere ontwerpstudies en analyses.

Toepassingsmogelijkheden

De wijktypen op de kaart kunnen worden gekoppeld aan mogelijke oplossingen voor klimaatbestendig inrichten. Een voorbeeld hiervan is het voorbeeldenboek over klimaatbestendig inrichten van woonstraten van de Hogeschool van Amsterdam (Kluck et al., 2017). Het voorbeeldenboek toont kostenbatenanalyses en varianten van waterbestendige inrichtingen. Voor acht verschillende wijktypen worden oplossingen gegeven en doorgerekend in een waterbalans en in kosten en baten.

De indeling in wijktypen biedt ook interessante mogelijkheden voor andere beleidsopgaven, zoals de energietransitie, elektrisch parkeren, vergroeningsstrategieën of integratie van sporten en spelen in de stad.

1.3. Aanpak

De wens voor de wijktypenkaart op de website van de Klimateffectatlas is om op basis van openbare data met een beperkte hoeveelheid handelingen te komen tot een duiding van de typologie van buurten in Nederland. De wijktypen worden onderscheiden op basis van eenvoudig beschikbare geografische data zoals bouwjaar, bouwhoogte, woningdichtheid, percentage groen en de functie van de bebouwing. De eerste stap is de bepaling van het wijktype op de kleine schaal van PC6 gebieden. Vervolgens worden die resultaten opgeschaald naar de PC5 en buurt niveaus. In de kleinere PC6 gebieden is de bebouwing vaak homogeen en is er duidelijk onderscheid te maken tussen wijktypen. Ga je naar een hoger schaal niveau dan worden de typen in een gebied meer gevarieerd. Op buurtniveau blijft de meerderheid van een typologie zichtbaar en blijft de kaart overzichtelijk voor heel Nederland.

De eerdere landelijke kaart met wijktypen was gebaseerd op een met de hand opgezette beslisboom, die destijds op basis van intelligente keuzes en kennis over wijktypen was opgesteld. Bijvoorbeeld een wijktype is tuinstad hoogbouw wanneer het bouwjaar tussen 1950 en 1965 ligt, de bebouwing tussen de vier en tien woonlagen heeft en er meer dan 35% groen is. De nieuwe kaart is ontwikkeld met gebruik van kunstmatige intelligentie. Het model heeft eerst regels ontwikkeld op basis van een validatieset. Vervolgens is voor alle individuele PC6 gebieden in de bebouwde kom van Nederland een wijktype bepaald.

1.4. Leeswijzer

Deze rapportage van het project beschrijft de genomen stappen en keuzes die zijn gemaakt om tot een verbeterde wijktypen kaart te komen.

In hoofdstuk 2 wordt de methodiek en toepassing van kunstmatige intelligentie beschreven. Daarna gaat hoofdstuk 3 in op de data die gebruikt is en hoe deze datasets klaar voor gebruik zijn gemaakt. In hoofdstuk 4 worden de resultaten van het computermodel gepresenteerd. De resultaten van de wijktypen worden op verschillende schaalniveaus weergegeven, in hoofdstuk 5 staat de uitwerking beschreven. In hoofdstuk 6 volgt een discussie met de belangrijkste verbeteringen en aanbevelingen voor verder onderzoek. Tot slot volgen de conclusies in hoofdstuk 7.

2. Methode bepalen wijktypen met behulp van kunstmatige intelligentie

Een computermodel kun je leren postcodegebieden onder te verdelen in wijktypen. Dit 'leren' gebeurt aan de hand van voorbeelden. Hoe meer voorbeelden het model heeft hoe nauwkeuriger de inschatting van nieuwe gebieden wordt. De invoer van deze voorbeelden vormt de validatieset die is gebaseerd op expertvalidatie. Paragraaf 2.1 beschrijft deze expertvalidatie. Daarna volgt een beschrijving van de inhoud van de data en het testen van de resultaten van de trainingset in paragraaf 2.2.

2.1. Methodiek training en toetsing aan de hand van expertvalidatie

Ter voorbereiding op het bepalen van de wijktypen voor heel Nederland zijn er eerst vijf testgemeenten gekozen. Binnen deze gemeenten hebben experts een selectie gemaakt van PC6 gebieden die samen de validatieset vormen. Voor deze geselecteerde PC6 gebieden hebben de experts het wijktipe bepaald op basis van visuele kenmerken en kenmerken zoals het bouwjaar.

De validatieset is opgedeeld in een trainingsset en een testset met een willekeurige verspreiding door de vijf testgemeenten. Het model 'leert' vanuit de trainingsset en 'meet' met de testset de nauwkeurigheid van de uitkomsten van het model (zie paragraaf 4.2 voor uitwerking). De trainingsset mag geen overlap hebben met de testset, daarom is gewerkt met een cross-validatie¹. De validatieset in zijn geheel bestaat uit bijna 10.000 PC6 gebieden. De verdeling over de wijktypes in [Tabel 1](#) geeft aan dat de historische binnenstad een groot aandeel heeft in de validatieset. Dat komt doordat deze typologie in alle vijf steden en dorpen het centrum vormt en eenvoudig toe te wijzen is binnen de contouren van het oude centrum.

Tabel 1: De dataset bevat 9.652 observaties met de volgende verdeling over de wijktypes

Historische binnenstad	2695
Naoorlogse woonwijk	813
Onbepaald	762
Tuinstad hoogbouw	702
Sub-urbane uitbreiding - VINEX	687
Volkswijk	628
Stedelijk bouwblok	607
Bloemkoolwijk	579
Tuinstad laagbouw	434
Tuindorp	414
Vernieuwd	298
Villa	267
Hoogbouw	261
Bedrijven	223
Groen	141
Vooroorlogse woonwijk	121
Buitengebied	11
Lintbebouwing	9

¹ Een cross-validatie betekent dat je bijvoorbeeld 10 train-test combinaties maakt (folds) waarbij de in de eerste fold de eerste 10% van de datapunten test is en de rest train, in de tweede fold is de tweede 10% test en de rest train etc. Ieder datapunt is dus 1x test en 9x train. Zo kan je een onafhankelijke voorspelling genereren voor ieder datapunt want deze wordt voorspeld door een model waarbij dat punt niet is meegenomen bij het trainen.



2.2. Data testen voor vijf testgemeenten

Voor het uitvoeren van het testmodel zijn de volgende gemeenten geselecteerd: Den Haag, Amsterdam, Utrecht, Deventer en gemeente Berkelland. De laatste gemeente bestaat uit de dorpen Borculo, Neede, Ruurlo en Eibergen. Deze vijf testgemeenten variëren in grootte en stedelijkheid. De variatie is belangrijk omdat bepaalde kenmerken die in een dorp anders zijn dan in een stad hierdoor beter worden aangeleerd door het model. Door het toevoegen van kleinere dorpen aan de validatieset kan het model een andere keus maken voor dorpen dan voor steden.

De validatieset die is geprepareerd als input voor het machine learning model bestaat uit één datapunt per PC6 gebied. Het datapunt bevat alle (steden)bouwkundige kenmerken (de X-variabelen) en het wijktype label (de y-variabele). Het wijktype label is toegekend door experts binnen het onderzoeksteam. Uit deze validatieset heeft het algoritme onderscheidende patronen leren herkennen. Deze zijn vervolgens geprojecteerd op de rest van Nederland om een voorspelling voor alle PC6 gebieden te geven.

De datapreparatie van de validatieset start met de lijst postcodes uit de vijf testgemeenten. Deze is vervolgens aangevuld door te koppelen met verschillende ruimtelijke datasets met specifieke kenmerken zoals bouwjaar, bouwhoogte en percentage groen. In sommige gevallen is de dataset niet in zijn geheel toegevoegd maar samengevat in één of twee waardes, om de complexiteit te beperken.

Omdat de validatieset niet alle mogelijke gevallen kan afdekken is ervoor gekozen om enkele beslisregels toe te passen om specifieke gevallen af te vangen. Zo is besloten om alle postcodes met een bouwhoogte boven de 30m naar het wijktype 'Hoogbouw' te forceren. Ook zijn er harde grenzen gezet op het toegestane bouwjaar van enkele wijktypes. Bijvoorbeeld: een Historische binnenstad is altijd gebouwd voor 1910 en een Stedelijk bouwblok komt niet meer voor na 1945. Daarnaast zijn de PC6 gebieden waarbij het model de categorie 'onbepaald' toewijst en waarvan de oppervlakte uit meer dan 70% groen bestaat geforceerd naar de classificatie groen.

De categorie Lintbebouwing wordt verwijderd uit de dataset. Dit is omdat Lintbebouwing maar 9 observaties in de data heeft en daarmee een te kleine categorie is, zie [Tabel 1](#). Deze buurten zullen voortaan in de categorieën Vooroorlogse woonwijk en Villa vallen. De categorie Onbepaald komt wel voor in de validatie set, maar niet in de trainingsset. Het is een relatief grote categorie, namelijk met 762 datapunten. Deze categorie is erg uiteenlopend, zonder eenduidig patroon wat deze groep kenmerkt. Bovendien is Onbepaald geen categorie die geleerd moet worden, maar eerder een restcategorie bij geen duidelijke herkenning. Daarom wordt het model getraind zonder de categorie Onbepaald. Omdat er uiteraard wel gebieden zullen zijn waarin de keuze voor een wijktype niet gemaakt kan worden, omdat de gebieden zeer gemengd zijn of niet bij een van de typologieën hoort die zijn gedefinieerd, zal een lage zekerheid in de keuze van het model toch leiden tot de categorie Onbepaald in de output.

3. Data voor wijktypen

Voor het bepalen van de wijktypen zijn verschillende datasets gebruikt die elk informatie bevatten over de gebouwde omgeving. In dit hoofdstuk lichten we de volgende aspecten toe: de gebruikte methodiek voor het verzamelen van de data, de verbeteringen en nieuwe toegevoegde datasets ten opzichte van de kaart uit 2017, datasets en een korte beschrijving van de geoprocessing van de verschillende datasets.

3.1. Datasets overzicht

Als input voor het model om de wijktypen voor heel Nederland te bepalen zijn verschillende datasets gebruikt. Een voorwaarde vanuit CAS was om gebruik te maken van openbare data. Ten opzichte van de 2017 versie zijn er nieuwe datasets toegevoegd, namelijk: dichtheid/stedelijkheid, de oriëntatie van straten, woningtype, type dak en het RUDIFUN-model van PBL. Daarnaast zijn ook verbeterde datasets toegevoegd, namelijk de Basisregistratie Adressen en Gebouwen (BAG), groenkaart en functie verdeling. Zie [Tabel 2](#) voor alle gebruikte datasets.

Als voorbereiding op de machine learning zijn alle datasets klaargemaakt in de GIS-programma's ArcGIS Pro en QGIS. Onderstaande paragraaf geeft een korte beschrijving van de geoprocessing. De geoprocessing van elke dataset staat stap voor stap beschreven en kan worden gevonden in de bijlage.

Voorafgaand aan de geoprocessing zijn de volgende twee beslissingen genomen: om de datasets met eenzelfde ruimtelijke verdeling te laten voorkomen:

- 1) Er is besloten om de grenzen van de PC6 gebieden gelijk te houden aan die van 2017. Op moment van verkenning van nieuwe datasets waren de PC6 gebieden met een nieuwe update enkel beschikbaar tegen betaling. Gezien de wens om enkel van openbare data gebruik te maken is besloten om de PC6 gebieden van 2017 te gebruiken. Een vergelijking met de PC6 gebieden in 2022 levert kleine verschillen op in grenzen.
- 2) Er is een bebouwde kom gedefinieerd die de ruimtelijke grens van bebouwd gebied weergeeft. Dit is gedaan omdat de buiten- en natuurgebieden niet mee dienden te worden genomen in de classificatie van wijk typologieën. Deze gebieden bevatten geen tot nauwelijks bebouwing en door deze gebieden niet mee te nemen gaat het definiëren van de wijktypen nauwkeuriger en sneller.

Tabel 2: Alle gebruikte datasets voor het produceren van de wijktype kaart met een omschrijving van de inhoud en keuze eenheid.

Dataset	Bron	Schaal	Inhoud	Keuze eenheid
Bouwjaar	3D BAG (2022)	Pand	De bouwjaar dataset is gebaseerd op de 3DBAG dataset. <i>Het bouwjaar dat wordt gehanteerd in deze dataset is het oorspronkelijke* bouwjaar (2018).</i> Deze dataset bevat het bouwjaar van elk geregistreerd object (Pand, Verblijfsobject, Ligplaats en Standplaats). *In de praktijk wordt bij renovatie en herontwikkeling van gebouwen het bouwjaar ingevoerd met de datum van de werkzaamheden.	Per postcode is het percentage van de bebouwing bekend per interval van 5 jaar. Bijvoorbeeld 25% van de bebouwing is van tussen 1925-1933 45% tussen 1970-1975 etc. De laagste categorie is '< 1890'. Uit deze dataset wordt de modus (meest voorkomende interval) gekozen en het minimum (jaartal van oudst aanwezige bouw). Het jongst aanwezige bouwjaar wordt niet gebruikt, omdat een pand gerenoveerd kan zijn. Dit zou patronen juist kunnen vertroebelen.
Bouwhoogte	3D BAG (2022)	Pand	De dataset met de gebouwhoogtes is gebaseerd op de 3DBAG dataset. Deze dataset bevat van elk geregistreerd object (Pand, Verblijfsobject, Ligplaats en Standplaats) de hoogte t.o.v. NAP van het maaiveld en het dak. Het dak niveau is berekend als het 70e percentiel van alle hoogtepunten op het corresponderende dak deel.	Per postcode is percentage van de bebouwing bekend per interval van 1 meter. De hoogste categorie is '> 30m'. Uit deze dataset is de modus gekozen.
Woningtype	Woningtype Esri (2022)	Pand	De dataset Woningtypering bevat het type woning per pand. Voor elk pand moet minimaal één verblijfsobject voorkomen met het gebruiksdoel "woonfunctie". Deze panden zijn met de analyse opgedeeld in de volgende categorieën: - Appartement. Dit is een woning waaraan meerdere verblijfsobjecten zijn gerelateerd, ongeacht het	Per postcode is percentage van bebouwing bekend voor de categorieën 'appartement', 'rijtjeshuis', '2-onder-1-kap' en 'vrijstaand'. De modus is gekozen.

			<p>gebruiksdoel van deze verblijfsobjecten. Echter, zoals voor alle panden in deze dataset, moet minimaal één van deze verblijfsobjecten het gebruiksdoel 'woonfunctie' hebben.</p> <ul style="list-style-type: none"> - Vrijstaande woning. Dit is een woning die niet aan een ander pand met een verblijfsobject is verbonden. - Tussen- of geschakelde woning. Deze woning is met meerdere panden met een verblijfsobject verbonden. In deze categorie vallen ook geschakelde woningen, dit zijn woningen die geschakeld zijn door middel van hun garages in plaats van hun hoofdgebouw. - Hoekwoning. Dit is de eerste of laatste woning in een serie van panden. Deze woning is zelf verbonden aan een enkel pand met een verblijfsobject, maar dit andere pand is verbonden met meerdere andere panden. - Twee-onder-een-kap. Deze woning is verbonden met een enkel pand met een verblijfsobject en dit andere pand is alleen met het eerstgenoemde pand verbonden. 	
Daktype	3D BAG (2022)	Pand	De data over het daktype komt van de 3D BAG dataset. In deze dataset is het daktype ingedeeld in 5 categorieën waar daktype 1 geen plat dak is, 3 gedeeltelijk plat dak en daktype 5 grotendeels plat dak.	Per postcode is percentage van het daktype bekend voor de categorieën 1 t/m 5. Hier heeft daktype 1 geen plat dak en daktype 5 grotendeels plat dak. De modus (meest voorkomende categorie) is gekozen.
Wegennetwerk	OSM (2021)	Buurt	De dataset is gebruikt voor de analyse van het aantal bearings en aantal intersecties. In deze dataset worden de kenmerken gebruikt die worden geïdentificeerd als primair, primary, primary_link, secondary, secondary_link, residential, tertiary, tertiary_link, motorway, trunk, trunk_link, living_street, motorway_link en unclassified.	<p>We hebben uit het wegennetwerk uiteindelijk 5 indicatoren meegenomen:</p> <p>'k_avg' (gemiddeld aantal wegen per kruispunt)</p> <p>'street_length_avg' (gemiddelde lengte van een wegsegment tussen twee kruispunten)</p> <p>'circuitry_avg' (maat voor de gekromdheid van wegen, namelijk de echte lengte gedeeld door de kortste rechte lijn tussen begin en eind)</p> <p>'dead_ends_per_node' (gemiddeld aantal doodlopende wegen per kruispunt)</p> <p>'bearings_top4' (variabele</p>

				ontworpen door onze 1e studentengroep die aangeeft welk aandeel van de wegen in één orientatierichting liggen (dus bij perfect vierkant wegennetwerk à la NY is dat 1).
Huishoudens per hectare	CBS (2022)	PC6	Deze dataset is gebaseerd op de huishoudens dataset van het CBS waarin huishoudens als volgt wordt gedefinieerd: Eén of meer personen die samen een woonruimte bewonen en zichzelf, dus niet-bedrijfsmatig, voorzien in de dagelijkse levensbehoeften.	Voor iedere postcode is het gemiddelde aantal huishoudens per hectare gekozen.
Percentage verharding	BGT (2022)		De Basisregistratie Grootchalige Topografie (BGT) bevat alle fysieke objecten in Nederland. Om de verharding te bepalen wordt gebruik gemaakt van de volgende lagen: wegdeel, tunneldeel, onbegroeidterreindeel, overbruggingsdeel, pand, gebouwinstallatie.	Voor iedere postcode is het gemiddelde percentage verharding toegevoegd.
Percentage wonen	BAG (2022)	Pand	De dataset met het percentage wonen is gebaseerd op de BAG dataset. In deze dataset wordt een categorisering (woon-, bijeenkomst-, cel-, gezondheidszorg-, industrie-, kantoor-, logies-, onderwijs, sport-, winkel-, overig gebruiksfunctie) aangebracht van het gebruiksdoel van een object	Voor iedere postcode is het percentage van het oppervlak met functie 'wonen' toegevoegd.
Percentage groen	Tauw (2023)	AHN4, BGT, NDVI	<p>1 – gras en lage vegetatie</p> <ul style="list-style-type: none"> • groenvoorziening • groenvoorziening: bodembedekkers • groenvoorziening: gras- en kruidachtigen • groenvoorziening: planten • grasland agrarisch • grasland overig • bouwland • bouwland: Akkerbouw • bouwland: Braakliggend • bouwland: Vollegrondsteelt • bouwland: Bollenteelt • kwelder • rietland • moeras <p>2 – struiken, heesters en middelhoge vegetatie</p> <ul style="list-style-type: none"> • groenvoorziening: heesters • groenvoorziening: struikrozen • heide • struiken • fruitteelt: klein fruit • fruitteelt: laagstam 	Voor iedere postcode is het totale percentage groen berekend en de percentages in de categorieën 'gras', 'heesters' en 'bomen'.

			<ul style="list-style-type: none"> • boomgaarden • fruitteelt: wijngaarden • fruitteelt • boomteelt 3 – bomen of bossen <ul style="list-style-type: none"> • gemengd bos • groenvoorziening: bosplantsoen • houtwal • loofbos • loofbos: griend en hakhout • naaldbos • fruitteelt: hoogstam boomgaarden 	
Percentage groen binnen erven	BGT (2022)		De Basisregistratie Grootchalige Topografie (BGT) bevat alle fysieke objecten in Nederland. Om het percentage groen te bepalen wordt gebruik gemaakt van de laag: erven, gecombineerd met de percentage groen dataset zoals hierboven beschreven.	Dezelfde procedure als 'percentage groen' maar hier wordt alleen het aandeel binnen de erfgronden meegeteld.
Industrie en bedrijven	TOP10NL, IBIS (2021)		De industriegebieden zijn afkomstig van de TOP10NL dataset. Deze dataset bevat alle industriegebieden in heel Nederland. Daarnaast is de dataset van IBIS toegevoegd waar alle geregistreerde bedrijventerreinen in meegenomen zijn.	Het percentage bebouwing met functie 'bedrijven' is toegevoegd.
Stedelijkheid	CBS (2022)	Buurt	Hierin is de indeling van gemeenten naar stedelijkheid gebaseerd op de gemiddelde omgevingsadressendichtheid (CBS, z.d.). Hierin zijn vijf categorieën te onderscheiden: <ul style="list-style-type: none"> - zeer sterk stedelijk met een gemiddelde van 2500 of meer adressen per km²; - sterk stedelijk: met een gemiddelde van 1500 tot 2500 adressen per km²; - matig stedelijk: met een gemiddelde van 1000 tot 1500 adressen per km²; - weinig stedelijk: met een gemiddelde van 500 tot 1000 adressen per km²; - niet stedelijk: met een gemiddelde van minder dan 500 adressen per km². 	Voor iedere postcode is de stedelijkheidsindex (op een schaal van 1-5) toegevoegd. Hier heeft categorie 1 een hoge mate van stedelijkheid en 5 laag.
Parken	TOP 10NL		De parken zijn afkomstig van de TOP10NL dataset. Deze dataset bevat alle parken in heel Nederland.	Voor iedere postcode is het percentage van het oppervlak wat binnen de grenzen van een park valt toegevoegd.
RUDIFUN	PBL		Deze dataset afkomstig van het PBL bevat informatie over dichtheid indicatoren per netto bouwblok. Uit	Deze dataset afkomstig van het PBL bevat informatie over dichtheid indicatoren per netto

			deze dataset wordt het gewogen gemiddelde van de Floor Space Index, Ground Space Index, Open Space Ratio, Gemiddeld aantal bouwlagen en de Mixed use Index met het bruto oppervlak gebruikt.	bouwblok. Uit deze dataset wordt het gewogen gemiddelde van de Floor Space Index, Ground Space Index, Open Space Ratio, Gemiddeld aantal bouwlagen en de Mixed use Index met het bruto oppervlak gebruikt.
--	--	--	--	--

3.2. Geoprocessing datasets

De geoprocessing is uitgevoerd in de GIS-programma's ArcGIS Pro en QGIS. Ook is er een nabewerking gedaan in het Microsoft-programma Excel. In ArcGIS hebben we gewerkt met de ModelBuilder. De ModelBuilder is een visuele programmeertaal voor het bouwen van geoprocessing modellen. Het voordeel is dat de bewerking van verschillende datasets geautomatiseerd kan worden, zowel de tussentijdse uitkomsten binnen het proces als de invoer voor een andere dataset.

De datasets zijn veelal omgezet van vectordata naar rasterdata om analyses te kunnen doen. De data voor heel Nederland is van grote omvang. De cell size van de rasters voor alle datasets met betrekking tot gebouwen is gezet op 2 bij 2 meter. Voor de datasets die informatie bevatten over landgebruik, zoals bijvoorbeeld de (on)verharding van het terrein, is gekozen voor een cell size van 1 bij 1 meter. Bij veel datasets is een Zonal Histogram uitgevoerd. Door middel van deze analyse kan de frequentieverdeling van waarden in de ene dataset binnen klassen van een andere dataset worden onderzocht. Met andere woorden, per PC6 gebied worden de voorkomende waardes van bijvoorbeeld bouwjaar, bouwhoogte, woningtype etc. berekend. In de bewerking van datasets met gebouwen is ervoor gekozen om enkel de footprint van de gebouwen in de PC6 gebieden te berekenen. De uitkomst van de Zonal Histogram geeft ook NODATA weer, wat staat voor het deel van de oppervlakte voor de PC6 gebieden waar geen gebouwen staan. Deze NODATA is geen kenmerk van een wijktype en daarom niet meegenomen in het berekenen van de percentages van voorkomende kenmerken van de bebouwing per PC6 gebied.

3.2.1. Bebouwde kom

De bebouwde kom dataset is gemaakt in samenwerking met CAS. De dataset is gebaseerd op de TOP10NL dataset en de CBS-stedelijkheid. Bij het maken van deze dataset was de TOP10NL (Plaatsen) uit 2019 de meest recente versie. In de TOP10NL dataset worden alle plaatsen weergegeven, maar de grenzen volgen niet de buurt of gemeentegrenzen. Op basis van de type gebieden van de plaatsen worden de relevante gebieden afgeleid. Uit de Top10NL plaats kaart zijn de 'Woonkern' en 'Industriekern' geselecteerd. Uit de CBS-dataset wordt het stedelijkheidsniveau (adressen per km²) per buurt meegenomen. Deze is geclassificeerd van 1 t/m 5 (zie [Tabel 2](#)). Vervolgens is gekeken naar de overlap (minstens 30%) tussen de geselecteerde gebieden uit de TOP10NL dataset en de buurten van de CBS-dataset. De uitkomst is de gehanteerde bebouwde kom dataset. Deze dataset geeft de ruimtelijke grens aan van de kaartlaag. Dat wil zeggen dat alle PC6 gebieden die binnen deze dataset vallen worden meegenomen in het definiëren van een wijktype.

3.2.2. Bouwjaar

De bouwjaar dataset is erg van belang in het bepalen van de wijktype, omdat bepaalde wijktypen in hele specifieke periodes zijn gebouwd, zoals de Naoorlogse Tuinstad (tussen 1950-1960). De dataset is gebaseerd op de 3D BAG dataset via Esri (2022). Deze dataset bevat alle panden die zijn geregistreerd, dus ook schuurtjes in tuinen. De schuurtjes geven een confuus beeld en helpen niet bij de herkenning van wijktypen. Daarom zijn deze schuurtjes verwijderd uit de dataset van de panden. Vervolgens zijn de overgebleven panden ingedeeld per interval van 5 jaar: <1890, 1890-1895, 1896-1900, 1901-1905, 1906-1910 2006-2010, 2011-2015, 2016-2020, >2020. Door het uitvoeren van een Zonal histogram is berekend hoeveel procent van bovengenoemde categorisering per PC6 voorkomt.

3.2.3. Bouwhoogte

De bouwhoogte van de gebouwen zegt veel over de wijktypen, zo zijn de woonlagen in een volkswijk lager dan bijvoorbeeld in een stedelijk bouwblok. De bouwhoogte dataset is gebaseerd op de 3D BAG dataset via ESRI (2022). De gebouwhoogte is berekend door de Ground 00 af te trekken van de Roof 75 hoogtewaarde. Alle gebouwen lager dan 2 meter zijn er uitgefilterd. De dataset bevat alle panden die zijn geregistreerd, dus ook schuurtjes in tuinen die niet relevant zijn voor het wijktypen project. Daarom is een bewerking uitgevoerd om de schuurtjes eruit te filteren.

Vervolgens is de data van de overgebleven panden opgedeeld met een interval van 1 meter vanaf 2 tot en met 30, na 30meter is er een categorie met >30. Door het uitvoeren van een Zonal histogram is berekend hoeveel procent van de bovengenoemde categorisering per PC6 gebied voorkomt.

3.2.4 Woningtype

De dataset Woningtype is gebaseerd op de Woningtypering dataset van Esri (2022). Deze dataset heeft de volgende categorieën: Appartement, hoekwoning, tussenwoning, 2-onder-1kap, vrijstaand.

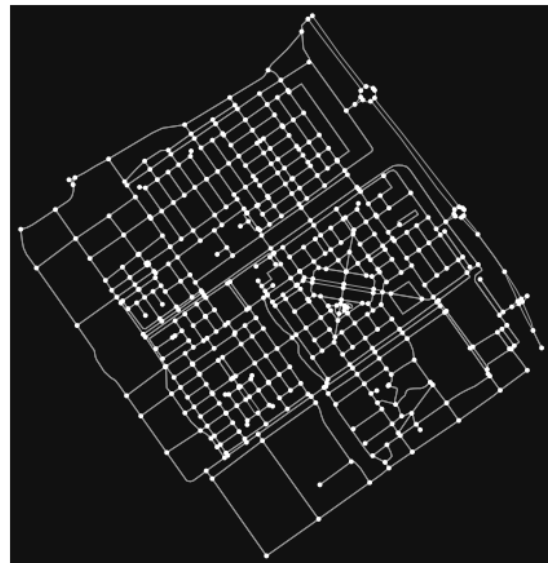
Er is besloten om de hoekwoning en tussenwoning samen te voegen tot rijtjeswoning. Vervolgens is er door middel van de Zonal histogram het percentage van het voorkomen van woningtypering per PC6 berekend.

3.2.5. Daktype

De daktype dataset is gebaseerd op de 3D BAG dataset via ESRI (2022). In deze dataset worden de daktypes aangegeven met cijfers van 1 t/m 5 waar: 1 = geen plat dak en 5 = grotendeels plat dak. Deze dataset bevat alle panden die zijn geregistreerd, dus ook schuurtjes in tuinen die niet relevant zijn voor het wijktypen project. Daarom is allereerst een bewerking uitgevoerd om de schuurtjes eruit te filteren. Van de overgebleven panden is door middel van een Zonal histogram het percentage van het voorkomen van daktype per PC6 gebied berekend.

3.2.6. Netwerkanalyse

Voor sommige wijktypen zijn niet de bouwkundige kenmerken zoals bouwjaar en bouwhoogte onderscheidend, maar juist het ontwerp van de wijk als geheel. Om dit in data te vatten is gekeken naar de eigenschappen van het wegennetwerk. Met name voor de Bloemkoolwijk is het wegennetwerk essentieel in het onderscheid tussen andere wijken uit dezelfde periode. Bloemkoolwijken hebben veel bochten en lussen, waar andere wijken meer kruisingen en rechte structuren hebben, zie [Figuur 4](#).



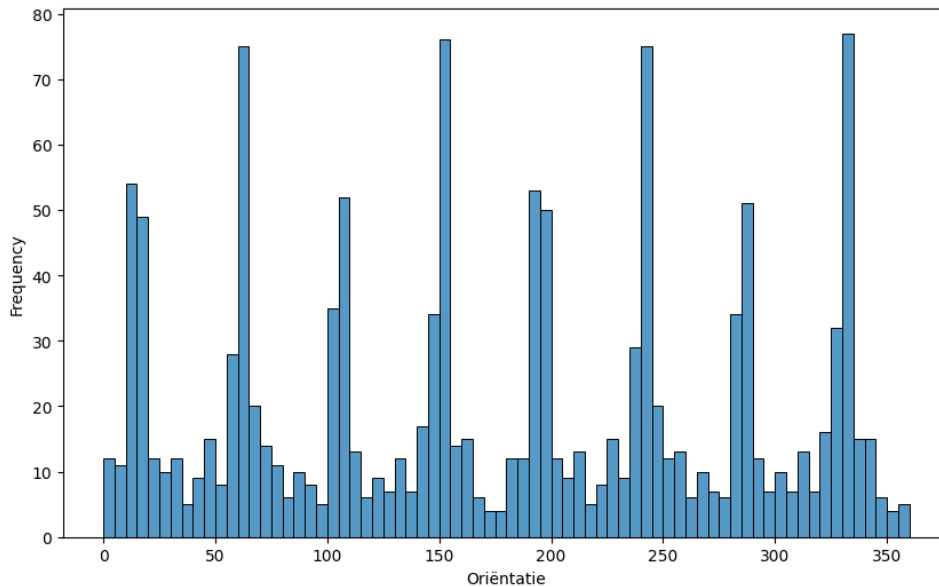
Figuur 4: Het wegenetwerk van een bloemkoolwijk (links) en een naoorlogse woonwijk (rechts) met dezelfde bouwperiode maar hele verschillende indeling.

De netwerkanalyse is gedaan met behulp van de gegevens van het Wegennetwerk uit de Open Street Maps dataset. De gegevens zijn beschikbaar per provincie en zijn met behulp van GIS samengevoegd. Binnen de dataset is een selectie van kenmerken gebruikt met de volgende attributen: *primary*, *primary_link*, *secondary*, *secondary_link*, *residential*, *tertiary*, *tertiary_link*, *motorway*, *trunk*, *trunk_link*, *living_street*, *motorway_link*, *unclassified*. Vervolgens zijn de intersecties bepaald met de geoprocessen *Unsplit* en *Pairwise*. Voor iedere postcode hebben we een analyse gedaan van het wegenetwerk in de omliggende buurt. Hieruit zijn enkele beschrijvende statistieken berekend zoals de dichtheid van kruispunten en straten, de verhouding straten/kruispunten en ook enkele zelf gedefinieerde getallen zoals de evenwijdigheid van het stratennetwerk, centraliteitsmaten van het netwerk en het aandeel doodlopende straten.

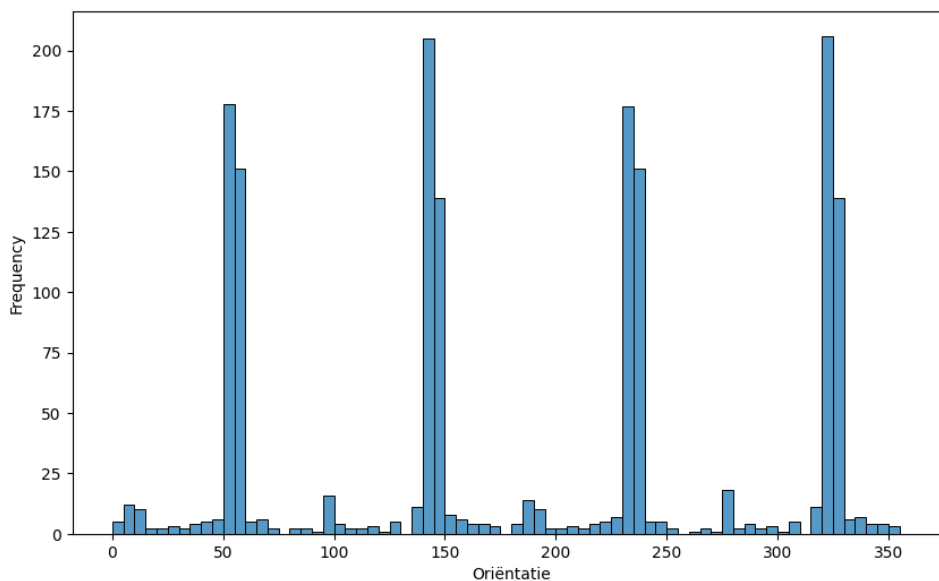
Omdat het wegenetwerk van een postcode-6 gebied weinig structuur bevat is deze analyse op buurtniveau uitgevoerd. Ieder PC6 gebied krijgt de netwerk-waarde van de buurt waarin het PC6 gebied zich bevindt. Voor iedere buurt is een analyse gedaan van het wegenetwerk waarin verschillende type wegen zijn geselecteerd waarin de auto, fiets en wandeloptyen worden meegenomen. Uit het wegenetwerk zijn de volgende statistieken berekend, gebruik makend van de python pakketten *networkx* en *osmnx*:

- *m* – aantal randen binnen de uitsnede
- *n* – aantal knooppunten binnen de uitsnede
- *k_avg* – gemiddelde oriëntatie knooppunten (graden in en graden uit)
- *edge_length_total* – straatlengte
- *edge_length_avg* - *edge_length_total* / *m* – gemiddelde straatlengte van de uitsnede
- *streets_per_node_avg* – aantal fysieke straten die verbonden zijn met elk knooppunt
- *street_segment_count* – aantal straatsegmenten in het netwerk
- *street_length_avg* – totale straatlengte / aantal straatsegmenten
- *circuity_avg* – gemiddelde gekromdheid van straten
- *self_loop_proportion* – percentage lussen
- *node_density_km* – knooppunten per km²
- *street_density_km* – straatlengte per km²
- *intersection_density_km* – aantal intersecties per km²
- *edge_density_km* – totale straatlengte per km²
- *area* - oppervlak buurt in km²
- *degree* – median of the degree centrality of each node
- *between* – mediaan van de betweenness centrality van elk knooppunt
- *closeness* – mediaan van de closeness centrality van elk knooppunt
- *triangles* – aantal driehoekconfiguraties in het netwerk
- *dead_ends* – aantal knooppunten met één graad
- *bearings_top4* – maat voor de piek van de verdeling van randoriëntaties (bearings)

De berekende statistieken van het netwerk kunnen helpen de verschillende wijktypes te onderscheiden. [Figuur 5](#) en [Figuur 6](#) laat twee histogrammen zien van de oriëntaties van de wegsegmenten (in graden) van bloemkoolwijken en naoorlogse woonwijken en bijbehorende verdeling van de oriëntaties (*bearings_top4*). Te zien is dat de naoorlogse woonwijk een hele sterke voorkeursoriëntatie heeft voor de wegen met 90 graden tussen de pieken, terwijl de bloemkoolwijk een meer continue verdeling heeft met een tweede set pieken op 45 graden. Hoe hoger het aandeel van wegen in een van de vier hoofdoriëntaties hoe rechthoekiger het wegenetwerk is.



Figuur 5: Aandeel van straatlengten in het netwerk wat voorkomt in de 4 meest voorkomende oriëntaties: 29% van de wegen in Bloemkoolwijken valt binnen een van de vier hoofd oriëntaties.



Figuur 6: Aandeel van straatlengten in het netwerk wat voorkomt in de 4 meest voorkomende oriëntaties: 83% van wegen in de Naorlogse woonwijken valt binnen een van de vier hoofd oriëntaties.

Door de netwerkanalyse met o.a. de oriëntatie van straten is de toekenning van het model voor Bloemkoolwijken versus Naorlogse woonwijk sterk verbeterd. Ook andere wijktypen krijgen een hogere score, zoals het onderscheid tussen Tuindorp en Volkswijk. De straatlengte van volkswijken is over het algemeen korter door het ontbreken van voortuinen en de minder diepe achtertuinten.

2.2.7. Huishoudens per hectare

De dichtheid van huishoudens wordt meegenomen omdat verwacht wordt dat de mate van dichtheid van huishoudens bepalend is voor het definiëren van de wijktipe. Bij bijvoorbeeld een villawijk zal het aantal huishoudens per hectare lager liggen dan in een bloemkoolwijk. De dataset huishoudens per hectare is gebaseerd op de CBS-dataset van 2022 en wordt aangeleverd op PC6 niveau. Omdat we gebruik maken van de PC6 dataset van 2017 moeten de PC6 gebieden van 2022 worden gekoppeld aan die van 2017. Door middel van

een Table join kunnen de gegevens uit 2022 gekoppeld worden aan de PC6 gebieden. Omdat de data vanuit het CBS wordt aangeleverd op PC6 niveau, is een omrekening van het aantal vierkante meters binnen de PC6 gebieden naar hectare uitgevoerd om antwoord te kunnen geven op het aantal huishoudens per hectare.

Vervolgens is een classificering aangebracht van aantal huishoudens per hectare: kleiner of gelijk aan 10 huishoudens per hectare, 10-20 huishoudens per hectare, 20-40 huishoudens per hectare, 40-100 huishoudens per hectare, 100-150 huishoudens per hectare en groter dan 150 huishoudens per hectare.

Als laatste stap is door middel van een Zonal histogram voor iedere postcode het gemiddelde aantal huishoudens per hectare toegevoegd.

3.2.8. Dichtheid/Stedelijkheid

De dichtheid van adressen is meegenomen om overeenkomsten te zien van wijktypen bij verschillende dichtheden. Een bloemkoolwijk in de randstad kan dichter bebouwd zijn dan een bloemkoolwijk buiten de randstad. Dit kan ook gelden voor bijvoorbeeld historische binnenstad die in een grotere stad een hogere dichtheid kan hebben dan in een kleine stad.

De stedelijkheid is gebaseerd op CBS-data. Hierin is de indeling van gemeenten naar stedelijkheid gebaseerd op de gemiddelde omgevingsadressendichtheid (CBS, z.d.), zie [Tabel 2](#) hierboven. Omdat deze data op buurtniveau beschikbaar is, is besloten om in elk PC6 gebied een middelpunt te berekenen. Dit middelpunt krijgt vervolgens een waarde toegewezen op basis van de ruimtelijke relatie die overeenkomt met de dataset van stedelijkheid van het CBS. De uitkomst is dat elk PC6 gebied een cijfer van de mate van stedelijkheid toegewezen heeft gekregen.

3.2.9. Verharding

De verdeling van verhard en onverhard is meegenomen in de wijktipe dataset omdat in een historische binnenstad vaak meer verharding op PC6 niveau is dan bij bijvoorbeeld een tuinstad of villawijk. Ook is de mate van verhard en onverhard belangrijk voor klimaatadaptatie mogelijkheden. De gebruikte data is afkomstig vanuit de BGT-dataset. Hierin zijn de volgende aspecten meegenomen om de mate van verharding te bepalen: Wegdeel, Tunneldeel, Onbegroeid terreindeel, Overbruggingsdeel, Pand, Gebouwinstallatie. Voor features met de fysieke vorm geclassificeerd als overgang, zand, halfverhard of onverhard is deze buiten de analyse gehouden. Houd er rekening mee dat de BGT-dataset geen onderscheid maakt tussen passerende klinkers en niet-doorlatende klinkers. Ook verharding binnen erven valt niet onder de dekking. Deze data is omgezet in rasterdata met een cell size van 2 bij 2 meter. Door middel van een Zonal histogram is het percentage van verharding per PC6 gebied berekend.

3.2.10 Wonen

Het percentage van panden die de functie 'wonen' hebben per PC6 wordt berekend om als eerste de typologie bedrijventerrein van de overige typologieën te onderscheiden. Daarnaast geeft de indeling van functies, anders dan wonen, mogelijk ook inzicht in het bepalen van historische binnensteden. In de historische binnenstad zijn panden vaak opgedeeld in functies zoals winkels en wonen.

Om het percentage wonen te berekenen is de dataset van BAG verblijfsobject gebruikt. Deze data wordt weergegeven in punten. Het model kan verschillende verblijffuncties over meerdere verdiepingen meenemen doordat het percentage van elke functie wordt berekend door de puntdata te linken aan de footprint van het gebouw. Dit wordt gedaan door een 'Spatial Join' en vervolgens een 'Add Join' om de oppervlaktes per functie te berekenen. Vervolgens wordt een onderscheid gemaakt tussen de functie wonen en niet wonen. Door middel van een 'Zonal histogram' is het percentage wonen per PC6 gebied berekend.

3.2.11. Groen

Sommige Wijktypen zijn van nature groener dan andere en daarom van groot belang voor een betere wijktypebepaling. Voor de analyse van het percentage groen per PC6 gebied is een groen kaart ontwikkeld door Tauw B.V. met input en feedback van de Hogeschool van Amsterdam. Tauw B.V. heeft de dataset geproduceerd met behulp van de software Tygron. De benodigde gegevens om de groenkaart te genereren zijn onder meer: AHN4, BGT en NDVI. Deze laatste dataset is ontwikkeld voor de wijktypen-update. De groenkaart is aangeleverd als een rasterdataset voor heel Nederland met een resolutie van 1 meter. De dataset bestond uit vier waarden, geen groen (0), gras (1), heesters (2) en bomen (3). Voor meer uitleg over de totstandkoming van deze dataset kunt u contact opnemen met Tauw B.V.

Om uit de rasterdataset het percentage groen per PC6 gebied te bepalen, is een 'Zonal histogram' gemaakt tussen de groenkaart en de PC6 gebieden binnen de bouwde kom. Nabewerking gebeurde in Excel om het percentage groen per PC6 gebied te bepalen.

3.2.12. Rudifun

Het PBL heeft een methode ontwikkeld om ruimtelijke dichtheden geautomatiseerd, landsdekkend en op een generieke wijze te berekenen. Deze dichtheden uitgedrukt in Floor Space Index (FSI) en Ground Space Index (GSI) en hun afgeleiden Layers (L) en Open Space Ratio (OSR). Deze indexen worden, op de schaalniveaus van bouwblok, buurt, wijk en gemeente aangeboden in de RUDIFUN-dataset (PBL, 2022). Uit deze dataset wordt het gewogen gemiddelde van de FSI, GSI, OSR, Gemiddeld aantal bouwlagen en de Mixed use Index met het bruto oppervlak gebruikt. Hier heeft verder geen geoprocessing plaatsgevonden, maar deze zijn direct ingevoerd in het model van kunstmatige intelligentie.

Deze dichtheidsindicatoren zijn een aanvulling op twee andere dichtheidsindicatoren: huishoudens per hectare (HH/ha) en stedelijkheid. Ze geven extra detailinformatie waar stedelijkheid maar vijf classes heeft en HH/ha alleen een dichtheid van aantal woningen. De aanvulling met de verzameling indicatoren in de RUDIFUN-dataset is de ruimtelijke component in oppervlakte en verdeling hiervan op het maaiveld en in de hoogte.

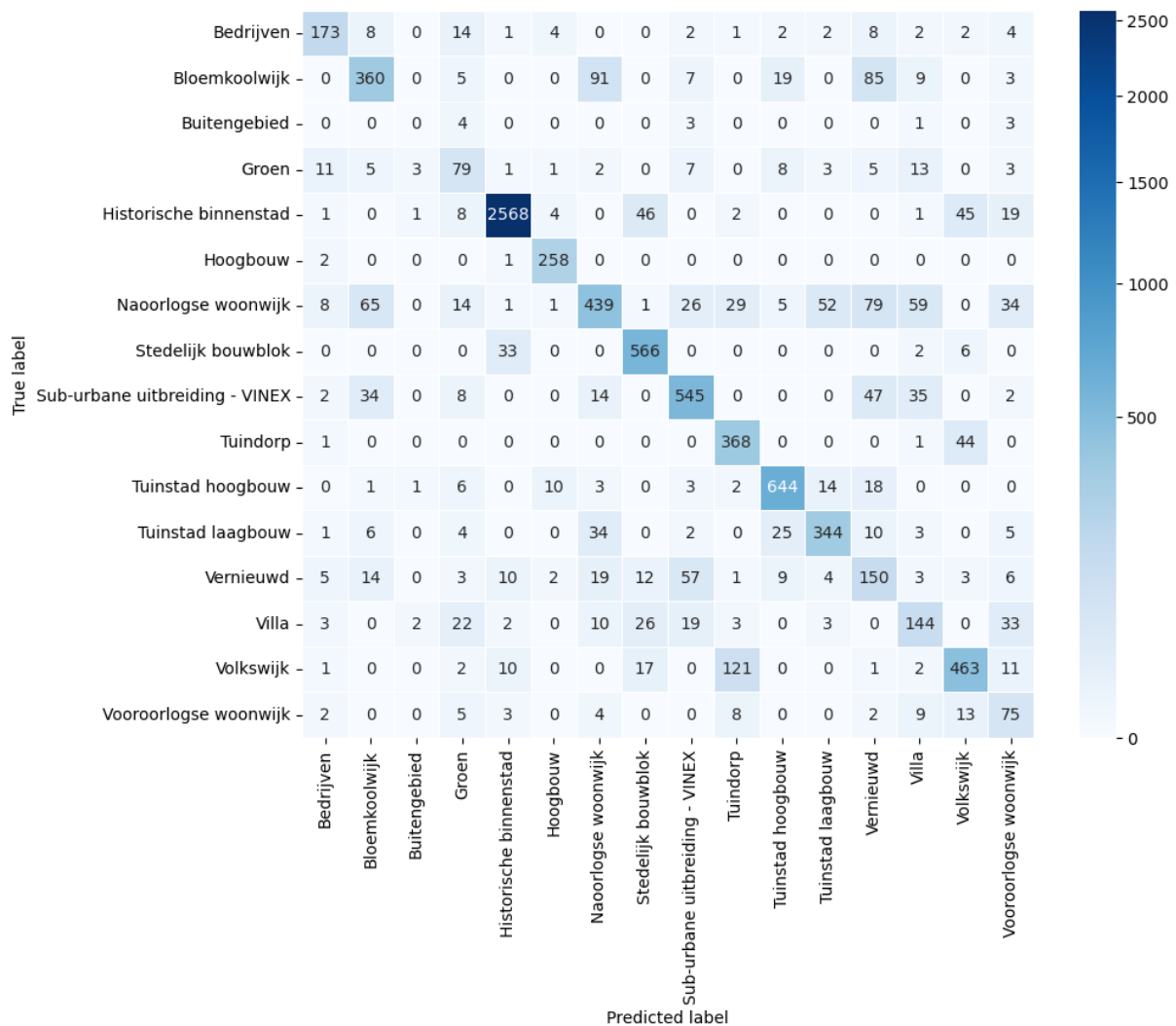
4. Machine learning inzetten voor het bepalen van wijktypen

Een voorspellend computermodel is getraind op basis van datasets (beschreven in hoofdstuk 3) met de invoer van wijktypen in de trainingsset (beschreven in hoofdstuk 2). Het getrainde model is ingezet om op basis van dezelfde datasets voor alle postcodes in Nederland het wijktype in te schatten.

4.1. Machine Learning modellering met het Random Forest Model

Nadat de data preparatie voltooid is, kan het model worden getraind met de trainingsset. Er is gekozen voor een Random Forest model (Ho 1995) omdat deze erg goed is in het meenemen van interactie-effecten. Dit houdt in: Het vinden van combinaties van verschillende variabelen die typisch zijn voor een bepaalde categorie. Daarnaast is het model goed in het omgaan met non-lineaire effecten. Dit zijn plotselinge veranderingen, zoals bijvoorbeeld het verdwijnen van een bouwstijl binnen vijf jaar.

Het op de trainingsset gebaseerde model wordt gebruikt om voor de testset wijktypen te bepalen. De Confusion Matrix in Figuur 7 toont de vergelijking van de bepaling door het model en de expertvalidatie voor de validatieset. In deze matrix is zichtbaar welke wijktypen nog vaak niet goed worden herkend. Zo worden er 2547 PC6 gebieden met het wijktype historische binnenstad goed voorspeld. Daarnaast krijgen 46 gebieden onterecht het label volkswijk, terwijl ze eigenlijk historische binnenstad zouden moeten zijn. De gevallen waar vaak een verkeerd label is gegeven zijn nader bekeken. Er bepaald of het toevoegen van extra datasets of aanpassen van huidige data beslissingen een verbeterde inschatting zou geven. Een Confusion Matrix is in het proces meerdere keren uitgevoerd. Dit heeft ertoe geleid dat er extra datasets zijn toegevoegd en er meer PC6 gebieden van een type aan de validatieset zijn toegevoegd. Hierdoor heeft het model meer voorbeelden om van te 'leren'.



Figuur 7: De Confusion Matrix voor het wijktypen model

Om de juiste parameters in te stellen voor het model is de data in subsets ingedeeld. Dit proces wordt hyperparameter tuning genoemd. Het model leert patronen uit een deel van de subsets en evalueert de kwaliteit op het andere deel. De prestaties worden vervolgens op nieuwe data gesimuleerd. Dat is belangrijk omdat we te maken hebben met ruimtelijke data waar autocorrelatie kan voorkomen. Wanneer we de subsets willekeurig zouden indelen, zoals gebruikelijk is, lopen we het risico dat het model leert van een postcodegebied en geëvalueerd wordt op zijn zwaar gecorreleerde buurgedeelte. Dit geeft een vertekend beeld van de kwaliteit van het model. De datapunten zijn dan niet onafhankelijk. Om dit effect tegen te gaan is ervoor gekozen om de subsets handmatig in te delen op buurtniveau. Zo zijn aangrenzende postcodegebieden altijd groepsgewijs in een subset ingedeeld en hierdoor is de evaluatie wel representatief.

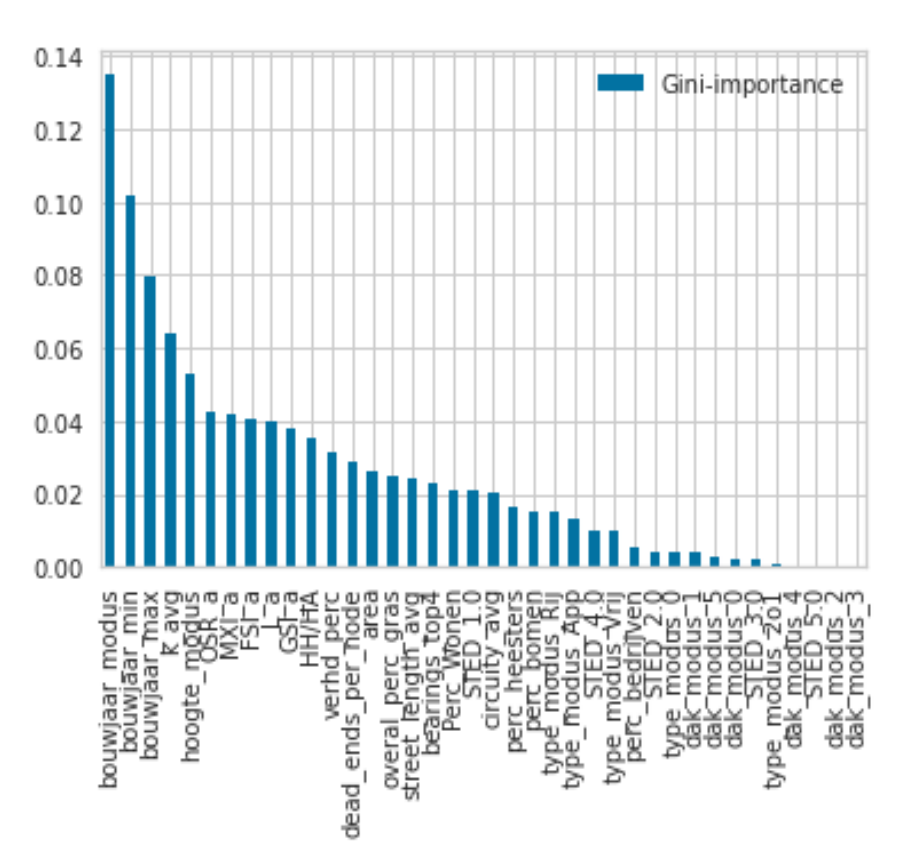
Het aantal observaties per typologie in de dataset zijn niet gelijk. Bijvoorbeeld hebben de categorieën Historische binnenstad en Villa respectievelijk 2695 en 267 observaties, zie ook Tabel 1 in Paragraaf 2.1. Om te zorgen dat het model niet alleen de kenmerken van de grotere typologieën leert, is de data gebalanceerd. Dit is gedaan door gewichten aan de datapunten te hangen, proportioneel aan de frequentie van de typologie.

Vervolgens is een systematische scan van combinaties van parameters om het best presterende model te vinden door telkens te trainen en te evalueren op de verschillende subsets. Dit wordt ook wel een grid search

genoemd. Hier zijn de volgende instellingen gevonden: Max dept= 15, n_estimators=100, min_samples_split=10, min_samples_lead=62. Deze zijn ook gebruikt voor het definitieve model.

4.2 Relevantie van de datasets

Wanneer het model getraind is, kunnen de meest invloedrijke variabelen worden bepaald (invloedrijk betekent hier dat de variabele vaak hoog in de beslisbomen voorkomt). Hieruit blijkt dat de variabele bouwjaar modus, gevolgd door bouwjaar_min en daarna hoogte_modus het meest invloedrijk zijn in de onderscheiding van de wijktypes. Dit is niet geheel onverwachts, de experts namen ook aan dat bouwjaar erg bepalend is voor het definiëren van een wijkttype. Ook is te zien dat de RUDIFUN-index van het PBL die de ruimtelijke dichtheden weergeeft van belang is voor het model om wijktypen te herkennen. Ook dit is niet verrassend omdat deze dataset inzicht geeft in de fysiek morfologische eigenschappen van de bebouwde omgeving. De variabelen die daarna van belang zijn is het percentage verhard, het percentage van type groen en het straatpatroon. De mate van stedelijkheid, type dak en woningtype hebben minder invloed in het bepalen van de wijktypen. Toch zijn deze van groot belang voor de herkenning van specifieke wijktypen zoals Villa.



Figuur 8: Relevantie van de data sets voor de bepaling van wijktypen door het Random Forest model (Gini-importance)

4.3. Confusion Matrix voor de wijkttype bepaling

In de Confusion Matrix is op de y-as en x-as respectievelijk de daadwerkelijke en voorspelde categorie weergegeven. Historische binnenstad is verreweg de meest voorspelde categorie, dit is verklaarbaar doordat dit ook de grootste categorie in de dataset was. Opvallend is dat de categorie Naoorlogse woonwijkrelatief

² Max depth = het maximaal aantal opeenvolgende splitsingen in een beslisboom, ook wel de diepte van de boom genoemd.
n_estimators het aantal beslisbomen in de random forest
min_samples_split: het minimale aantal datapunten wat nog gesplitst mag worden
min_samples_leaf: het minimale aantal datapunten wat een eindpunt (blad van een boom) mag overblijven

vaak verkeerd voorspeld wordt. Daarnaast is ook opvallend dat Volkswijken regelmatig voorspeld worden als Tuindorpen.

Het model rekent vervolgens voor elk postcodegebied een kans uit voor ieder mogelijk wijktype. Dit telt op tot 100%. Het model classificeert het postcodegebied als het wijktype corresponderend met de hoogste kans. Als we deze voorspellingen vergelijken met de expertclassificaties komen we op een nauwkeurigheid van 82%. Figuur 7 toont de uiteindelijke Confusion Matrix.

Het model geeft per PC6 aan met welke zekerheid de wijktype voorspelling is gedaan. Wanneer het model de voorspelling met 20% zekerheid of minder doet wordt het wijktype omgezet naar 'Onbepaald'. Zo komen er minder twijfelachtige uitkomsten door in de kaart. Dit heeft uiteraard invloed op de nauwkeurigheid, wanneer we namelijk al deze 'onbepaalde' PC6 gebieden buiten beschouwing laten wordt de nauwkeurigheid van de wijktypebepaling 85%.

4.3. Nauwkeurigheid bepaling wijktypen door het model

De accuraatheid kan soms een vertekend beeld geven wanneer er een hoge klasse-onbalans aanwezig is. Dit houdt in dat een veel voorkomende categorie vaak correct voorspeld wordt, maar de kleinere categorieën vaak minder correct voorspeld worden. De accuraatheid blijft dan steeds hoog, terwijl er veel wijktypen fout worden voorspeld. Daarom is naast de accuraatheid ook de precision, recall en f1-score berekend per categorie in Tabel 3.

Precision is de ratio tussen hoe vaak het wijktype correct is voorspeld (True Positives - tp) en hoe vaak het wijktype in totaal is voorspeld. In de confusion matrix in staat de precision op de horizontale as van de grafiek. Wanneer deze waarde laag is, betekent dit dat het model deze categorie vaak voorspeld heeft, maar deze vaak niet klopt. Een zeer conservatief model, wat inhoudt dat je alleen iets voorspelt bij een zeer hoge zekerheid, zal een hoge precision halen. Er worden dan weinig datapunten gevonden, maar degenen die gevonden worden kloppen wel.

Recall is de ratio tussen hoe vaak het wijktype correct is voorspeld (tp) en hoeveel datapunten dit wijktype daadwerkelijk heeft. Dit geeft een indicatie hoeveel van de datapunten door het model gevonden zijn. In de confusion matrix in staat de recall op de verticale as van de grafiek. Wanneer deze waarde laag is, betekent dit dat het wijktype vaak voorkomt, maar vaak niet als dit wijktype wordt voorspeld. Een zeer laagdrempelig model, wat al een voorspelling geeft bij een lage zekerheid, zal een hoge recall halen want alle datapunten worden gevonden ten kosten van veel 'false positives'.

De f1-score is het harmonisch gemiddelde van de precision en recall balanceert het conservatieve en het laagdrempelige om tot een evenwichtige inschatting te komen:

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2tp}{2tp + fp + fn}$$

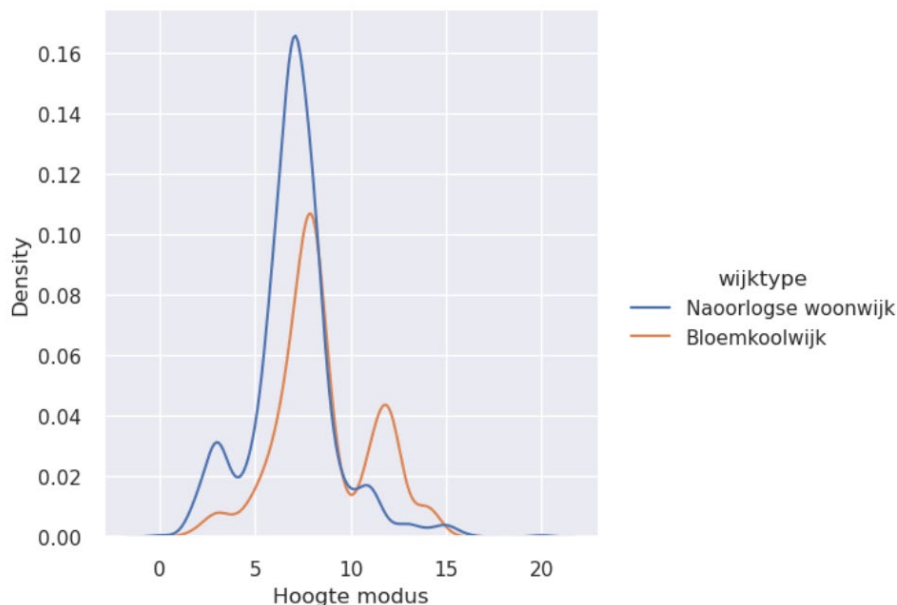
Tabel 3: De precision, recall en F1-score geven een meer gebalanceerd beeld van de behaalde nauwkeurigheid door het model

	precision	recall	f1-score	support
Bedrijven	0.82	0.77	0.79	223
Bloemkoolwijk	0.82	0.75	0.79	579
Buitengebied	0.00	0.00	0.00	11
Groen	0.48	0.55	0.51	141
Historische binnenstad	0.98	0.95	0.96	2695
Hoogbouw	0.92	0.97	0.95	261
Naoorlogse woonwijk	0.68	0.50	0.58	813

Stedelijk bouwblok	0.86	0.92	0.89	607
Sub-urbane uitbreiding - VINEX	0.78	0.83	0.80	687
Tuindorp	0.70	0.85	0.77	414
Tuinstad hoogbouw	0.92	0.92	0.92	702
Tuinstad laagbouw	0.79	0.80	0.80	434
Vernieuwd	0.49	0.51	0.50	298
Villa	0.47	0.56	0.51	267
Volkswijk	0.75	0.77	0.76	628
Vooroorlogse woonwijk	0.37	0.57	0.45	121
accuracy			0.81	8881
macro avg	0.68	0.70	0.69	8881

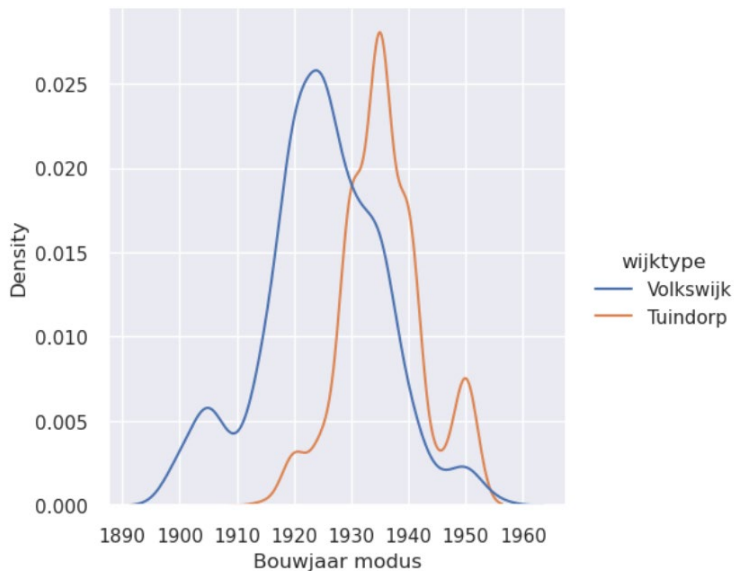
Groen en Vooroorlogse woonwijk scoren beiden op alle prestatie indicatoren relatief laag. Dit kan liggen aan dat de wijktypes weinig observaties bevatten in de validatieset. Het wijkttype Historische binnenstad scoort heel hoog, dit was te verwachten aangezien dit de grootste categorie van de validatiedataset is. Verder scoort hoogbouw hoog, terwijl het niet veel datapunten bevat. Dit ligt waarschijnlijk aan het feit dat hoogbouw een harde grens heeft, namelijk als de bouwhoogte hoger is dan 30 meter is het PC6 gebied standaard hoogbouw. Tevens heeft Naoorlogse woonwijk een hoge precision en een lage recall. Dit betekent dat Naoorlogse woonwijk veel datapunten heeft (813) maar niet vaak als Naoorlogse woonwijk wordt voorspeld (602). Dit is ook zichtbaar in de Confusion Matrix in . Daarnaast scoort ook het wijkttype Villa vrij laag. Dit kan liggen aan dat Villa uiteenlopende kenmerken heeft. Het model heeft moeite met het voorspellen van dit wijkttype. Dit is wederom terug te zien in de Confusion Matrix.

Datasets geven in combinatie met elkaar de mogelijkheid om onderscheid te maken tussen verschillende wijktypen. Wijktypen die veel op elkaar lijken in bouwhoogte zoals de bloemkoolwijken en naoorlogse woonwijken (zie Figuur 9), moeten daarom onderscheiden worden met behulp van andere factoren zoals een netwerkanalyse.

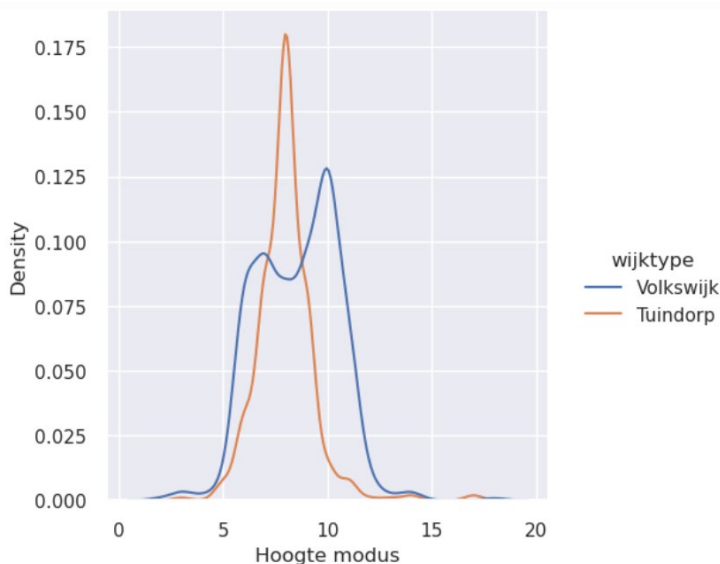


Figuur 9: Vergelijking van de Hoogte modus tussen Bloemkoolwijk versus Naoorlogse woonwijk

Een ander voorbeeld van wijken die moeilijk te onderscheiden zijn, zijn tuindorpen en volkswijken. Het bouwjaar overlapt elkaar (Figuur 10) maar er is wel meer kans dat het een tuindorp is na 1930. Er is daarnaast meer kans dat het een volkswijk is voor 1930. Het model kijkt ook naar de bouwhoogte (Figuur 11) van de twee wijktypen en 'leert' dat volkswijken vaker boven de 10meter hoogte komen en tuindorpen vrijwel niet. Bij een hoogte boven de 10 meter kan het model de keus voor volkswijk met meer zekerheid maken. Hoe meer onderscheidende databronnen, hoe beter het model kan voorspellen. Toch is een selectie in data nodig voor een goede balans tussen zoveel mogelijk onderscheidende data en de rekensnelheid en -capaciteit.

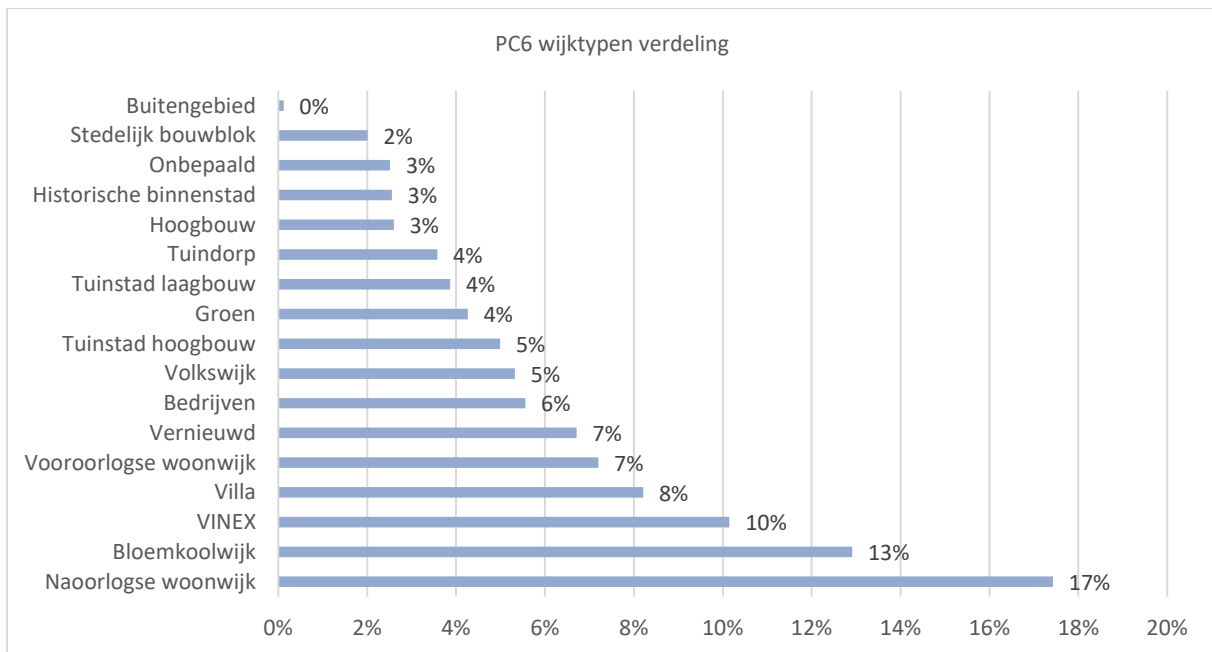


Figuur 10: Vergelijking van de Bouwjaar modus tussen Volkswijk en Tuindorp



Figuur 11: Vergelijking van de Hoogte modus tussen Volkswijk versus Tuindorp

Nadat het model is geoptimaliseerd wordt als laatste stap de gehele dataset op het model getraind en opgeslagen. Nu geconcludeerd is dat het model een goede voorspelling geeft voor de wijktypes, wordt het model gebruikt om een voorspelling van wijktypes van heel Nederland te maken. De verdeling is weergegeven in Figuur 12.



Figuur 12: Verdeling van PC6 wijktypen in percentage

De tabel toont dat de grootste groep (met iets minder dan 20%) Naoorlogse woonwijk is. Dit is goed te verklaren omdat deze categorie over een lange periode verspreid is, vanaf 1945 tot heden. Maar ook omdat deze typologie in alle steden en dorpen voorkomt, als uitbreidingswijken en in binnenstedelijke verdichting. De historische binnenstad komt in de praktijk minder vaak voor dan in de validatieset. De duidelijke contouren van een historische kern vormen snel veel PC6 gebieden waardoor deze makkelijker aan de handmatig ingevoerde validatieset zijn toe te voegen dan de andere gebieden waarin de toekenning van het wijkttype meestal veel meer tijd kost.

5. Clustering van PC6 gebieden

Zoals eerder beschreven zijn we in de aanpak begonnen met het duiden van de typologie op PC6 niveau omdat op dit niveau de bebouwing vaak homogeen is en daarmee duidelijk onderscheid te maken valt tussen wijktypen. Nadat we de resultaten van de PC6 gebieden hebben verkregen zijn de PC6 gebieden samengevoegd tot PC5 gebieden en vervolgens opgeschaald naar meest voorkomende typologie per buurt. Deze 2-staps duiding is nodig omdat veel buurten in de praktijk niet homogeen zijn, maar bestaan uit verschillende typologieën door inbreidingen en aanpassingen. Door gegevens samen te voegen op buurt niveau blijven de kaarten voor heel Nederland overzichtelijk.

Om op een hoger ruimtelijk niveau een schatting te maken van het wijkttype zijn de uitkomsten van de PC6 gebieden geaggregeerd naar PC5 gebieden en buurten. Hierbij is telkens het gewogen gemiddelde genomen van de uitkomsten van alle PC6 gebieden die omsloten werden door de grenzen van een PC5 gebied of buurt. Van deze omsloten PC6 gebieden is het meest voorkomende wijkttype bepaald, gewogen naar het omsloten oppervlak en de betrouwbaarheidsscore uit de vorige stap. Zo draagt een gebied met een zeer betrouwbare uitkomst relatief zwaarder mee aan de einduitkomst, maar ook een gebied met een groot oppervlak (alleen het gedeelte van het PC6 oppervlak wat daadwerkelijk binnen de grenzen ligt is hierbij meegenomen).

Uit deze berekening komt een wijkttype met de sterkste aanwezigheid binnen het gebied en opnieuw een betrouwbaarheidsgetal welke is gedefinieerd als de uitkomst van het gewogen gemiddelde van de winnaar ten opzichte van het totaal. Als er bijvoorbeeld drie wijktypes sterk aanwezig zijn in een buurt zal de hoogste waarde slechts een-derde vertegenwoordigen, maar in een zeer homogene buurt zal dit richting de 100% gaan.

De analyse is uitgevoerd voor heel Nederland met behulp van het python package geopandas. Zie Figuur 12 en 13 voor de verdeling van de wijktypen op PC5 en buurtniveau.

5.1. Beslisregels wijktypen toekennen per buurt.

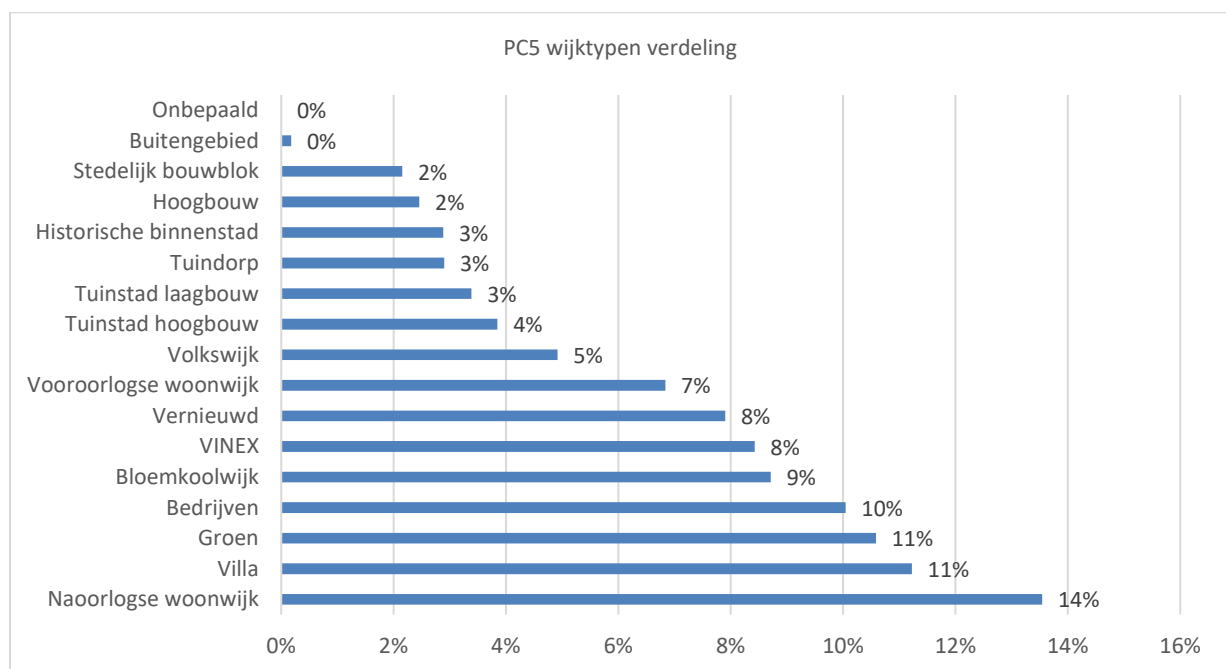
Voor buurten is prioriteit aan de woongebieden gegeven omdat de wijktypenkaart als basis dient voor het inzichtelijk maken van klimaatadaptatiemogelijkheden in bebouwd gebied. Daarom worden buurten met het wijktipe groen of bedrijven met een lage "confidence" waarde geforceerd naar de op een na hoogst scorend wijktipe, zolang het maar een woongebied is. De volgende prioriteringslogica is aangehouden voor het samenvoegen van PC6 naar PC5 en buurt:

Woonwijken

- Hoogste "confidence" percentage

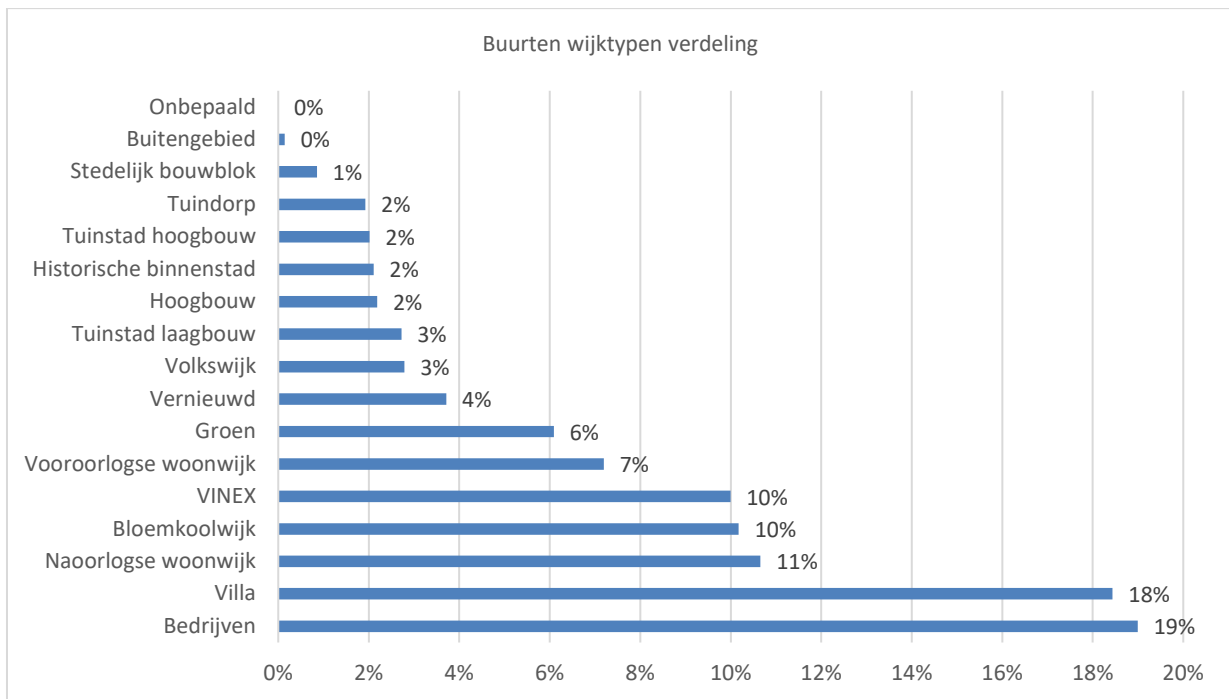
Groen en Bedrijven

- Als het hoogste "confidence" percentage $< 0,71$ is, kies dan het wijktipe met het op een na hoogste percentage "confidence"
- Als een na hoogste "confidence" percentage $< 0,20$ is, behouden label Groen of Bedrijven



Figuur 13: Verdeling van wijktypen op PC5-niveau in procenten.

In de verdeling van wijktypen op PC5 niveau blijft net als in de PC6 de Naoorlogse woonwijk de meest voorkomende. De Bloemkoolwijk en VINEX leggen het echter af tegen Villa, Groen en Bedrijven.



Figuur 14: Verdeling van wijktypen op buurtniveau in procenten.

In de verdeling van wijktypen op buurtniveau schuiven de typen Bedrijven en Villa naar de meest voorkomende typen. Een verklaring voor het verschil is dat deze typen binnen veel buurten aanwezig zijn en door hun grote oppervlak vaker de overhand krijgen dan een wijktypen met een kleiner oppervlak.

6. Conclusies

Door Machine Learning in te zetten bij het bepalen van wijktypen is een verbeterde kaart ontwikkeld. De kaart van 2017 was gebaseerd op een handgemaakte beslisboom en had beperkingen in het herkennen van ruimtelijke variaties. Om hierin verbetering te brengen is in deze studie onderzoek uitgevoerd waarbij kunstmatige intelligentie heeft bijgedragen aan het onderscheiden van wijktypen voor Nederland. De uiteindelijke nauwkeurigheid waarmee het model een wijktype voorspeld is gemiddeld 82%.

Verbetering Bloemkoolwijken en Naoorlogse woonwijken

De toevoeging van de netwerkanalyse heeft het onderscheid tussen bloemkoolwijken en naoorlogse woonwijken verbeterd. Het bouwjaar van deze wijken overlapt en ook de bouwhoogte. Met alleen die gegevens kan het model moeilijk een keuze maken. Met de netwerkanalyse kan het model de lussen en de oriëntatie van wegen in een buurt meenemen in de keuze. De netwerkanalyse is gedaan op buurtniveau waarna de uitkomst aan alle PC6 gebieden in die buurt is meegegeven.

Verbetering andere wijktypen

Een verbetering van het wijktype Villa is gedaan door het toevoegen van de dataset Woningtype. Een vrijstaand huis of 2 onder 1 kap zijn voor het model een indicatie dat een PC6 gebied Villa is. In combinatie met een hoog percentage groen en lage stedelijkheid wordt deze keuze duidelijker. De wijktypen Villa en VINEX liggen nog steeds dichtbij elkaar wanneer een VINEX wijk grote vrijstaande huizen op kavels van redelijk formaat bevat. Ook de wijktypen Villa en Vooroorlogse woonwijk liggen dichtbij elkaar wanneer aan de rand van dorpen zich vrijstaande huizen of voormalige boerderijen bevinden.

Een ander voorbeeld van een verbetering is het onderscheid tussen de wijktypen Tuindorp en Volkswijk. Deze hebben een overlap in bouwperiode, vergelijkbare bouwhoogte en stratenpatroon. Er zijn kleine nuances waarmee het model toch onderscheid kan maken. Volkstuinen hebben meer privaat groen en een iets hogere

bouwhoogte. Door de bouwhoogte een kleiner interval van 1 meter (tot 30 meter hoogte) in plaats van 3 meter kan het model een nauwkeuriger voorspelling doen.

Dorpen representatief

De belangrijkste verbeteringen van deze kaart ten opzichte van de 2017 versie zijn dat de wijktypen in dorpen meer representatief zijn en niet in hun geheel bloemkoolwijken zijn. Door het toevoegen van verbeterde en nieuwe datasets is meer actuele data gebruikt om het wijktypen op PC6 niveau te bepalen.

Extra detail niveau met kaartlaag op PC6 niveau

Door een 2-staps clustering van PC6 naar PC5 naar buurtniveau, is de clustering preciezer en zijn wijktypen op buurtniveau beter te verklaren door de gebruiker. De presentatie van PC6 gebieden en buurten is dus verbeterd ten opzichte van de 2017 kaart en hierdoor beter bruikbaar om inzicht te krijgen in de ruimtelijke kenmerken van een gebied. Tot slot zijn er meer PC6 gebieden toegevoegd.

Het model geeft een goede voorspelling van de wijktypen. Het eindproduct is de kaart met de gebouwde omgeving van Nederland, opgedeeld in wijktypen op PC6- en buurtniveau. Deze kaart kan goed worden gebruikt voor gemeenten en adviesbureaus om mogelijke klimaatadaptatie maatregelen te verkennen op basis van de wijktypologie.

De categorie 'onbepaald' voor beter overzicht

In het proces bleek soms dat een aanpassing ter verbetering van het ene wijktipe een minder goede voorspelling voor een ander wijktipe gaf. Niet alle bebouwing is simpelweg in te delen in de 16 gekozen classificaties. Met een uitbreiding in het aantal wijktypen kan de voorspelling van het model nauwkeuriger worden, maar dat worden dan steeds kleinere subgroepen. De PC6 gebieden die nu als 'onbepaald' worden geduid zijn ofwel een tussenvorm van meerdere wijktypen of komen weinig voor. Een uitbreiding van het aantal wijktypen zal het overzicht en herkenbaarheid van de typen niet ten goede komen.

Toepassing van de wijktypen in duiding klimaatrisico's en kansen

Voor alle wijktypen is een nieuwe inschatting gemaakt van de klimaatrisico's wateroverlast en hitte. Die voor hitte zijn bepaald op basis van een analyse van de gevoelstemperatuurkaart door Deltares en de HVA, zie ook Bijlage 1. Die voor water zijn overgenomen uit de vorige tabel uit 2017 waarvoor een analyse van wateroverlast was gedaan. Met deze tabel is direct duidelijk of er in een wijk veel of weinig aandacht voor een klimaatrisico moet zijn.

Tabel 4: Kenmerkentabel wijktypen met bouwperiode, aantal bouwlagen, percentage groen en een indicatie voor kwetsbaarheid hittestress en wateroverlast op buurtniveau.

Wijktipe	Bouwperiode	Hoogte in lagen	Percentage groen	Kwetsbaarheid hitte	Kwetsbaarheid water
Historische binnenstad	Voor 1910	3 tot 5	2-13%	+++	+++
Vooroorlogse woonwijk	Voor 1945	2 tot 4	16-37%	+	++
Stedelijk bouwblok	1890-1940	4 tot 6	6-18%	+++	+++
Tuindorp	1910-1940	2 tot 3	9-26%	+++	++
Volkswijk	1910-1940	2 tot 3	4-15%	+++	+++
Naoorlogse woonwijk	Na 1945	2 tot 3	12-29%	++	++
Tuinstad laagbouw	1945-1960	2 tot 3	11-29%	++	++
Tuinstad hoogbouw	1950-1965	4 tot 10	18-46%	++	++
Hoogbouw	Na 1960	6 tot 20	4-46%	+	++
Bloemkoolwijk	1970-1990	3 tot 5	15-36%	++	+
Sub-urbane uitbreiding - VINEX	1990-2020	3 tot 10	13-32%	++	+
Villa	van alle tijden	1 tot 4	36-58%	+	+

Vernieuwd	van alle tijden	2 tot 6	4-21%	+++	++
Bedrijven	van alle tijden	2 tot 3	9-35%	+	+

Specifieke kansen in wijktypen kunnen ook in beeld worden gebracht door kenmerken als percentage schaduw, infiltratie capaciteit of aanduiding grondwaterniveau.

7. Discussie

De huidige Machine Learning methode kent naast vele voordelen ook beperkingen. In de volgende paragrafen geven we een overzicht van de beperkingen van het gekozen model, de beschikbare data, de expert validatie en de mogelijke verbeteringen.

7.1. Beperkingen van methodiek

De beperkingen van de huidige methodiek bespreken we hieronder per onderdeel; de expert validatie, grenzen van de PC6 gebieden en buurten, gemixte buurten, vernieuwing en bouwjaar.

Expert validatie

De toewijzing van wijktypen in de validatieset, met een grote hoeveelheid van ruim 9.000 PC6 gebieden is handmatig toegekend door drie experts. Van tevoren is afgestemd wanneer een wijktype tot een specifiek wijktype behoort en wanneer niet. Bij twijfel is het wijktype voor dat PC6 gebied niet bepaald. Ondanks deze aanpak is er nog steeds de mogelijkheid van een verschil in wijktypen tussen de experts. Het model zal herkennen dat er in sommige wijktypen overlap zit en dat meenemen in de automatische keuze door een minder grote zekerheid mee te geven.

Met de variatie in stedelijkheid en voorkomen van wijktypen tussen de 5 test gemeenten is getracht van alle wijktypen voldoende PC6 gebieden in de validatieset op te nemen. Desondanks is de validatieset niet helemaal evenwichtig geworden. Zo is er relatief veel historische binnenstad aangewezen en zijn er minder gebieden zoals villawijken en vooroorlogse woonwijken gedefinieerd. Een aantal wijktypen komen minder vaak voor, maar zijn wel relevant voor ofwel het beter indelen van dorpen, ofwel het uiteindelijk inschatten van kwetsbaarheid voor klimaatverandering en het toepassen van adaptatiemaatregelen. De onevenwichtige verdeling heeft beperkingen gehad in het toewijzen van een typologie door het model, omdat het model niet voor alle wijktypen evengoed kon 'leren' welke ruimtelijke aspecten van belang zijn.

Grenzen van PC6 gebieden en buurten

De grenzen van PC6 gebieden en buurten zijn niet overal logisch verdeeld ten opzichte van de aanwezige bebouwing. Zo kan het voorkomen dat een PC6 gebied een groot oppervlak beslaat of uit meerdere polygonen (deelgebieden die niet aaneengesloten zijn) bestaat en daardoor geen homogene bebouwing bevat. Daarnaast kunnen er ook nog grote delen onbebouwd terrein aanwezig zijn. Met andere woorden, hierdoor kunnen verschillende wijktypen bij elkaar zitten. Ook delen van een park, groen of onbebouwd kunnen onderdeel zijn van een gebied. Met de toekenning van een wijktype in een gemixt PC6 gebied kan het zijn dat delen van een park bebouwd lijken of juist dat delen van bebouwde wijken gemiddeld een heel hoog percentage groen hebben. In de GIS-analyse is daarom voor elke polygoon binnen een PC6 gebied een eigen ID gecreëerd met een eigen wijktype.

Een mix van bebouwing blijft een mix

Het blijft lastig om aan gemixte gebieden (op PC6 of buurtenniveau) één enkel wijktype toe te kennen. Dit ligt niet zozeer aan een beperking in de bepaling of aan gebrek aan goede data, maar simpelweg aan het feit dat gebieden soms een mix zijn.

Vernieuwing met nieuw bouwjaar

Door inbreiding of vernieuwingen in een wijk (zoals nieuwbouw in oude stijl) is het lastig om specifieke wijktypen te onderscheiden. Dit is extra lastig omdat het bouwjaar, waar het model veel keuzes in typologieën typeert, niet het oorspronkelijke bouwjaar blijft behouden na renovatie. Een volkswijk uit 1925 die grootschalig wordt gerenoveerd krijgt hierdoor het bouwjaar van bijvoorbeeld 1985. Het model maakt dan de keuze om de wijk toe te kennen aan de typologie naoorlogse woonwijk. Wat wel goed gaat zijn gesloopte panden die als wijktipe vernieuwd terugkomen. Het model herkent een verschil in bouwperiode van minimaal 20 jaar tussen een PC6 gebied en de omringende PC6 gebieden.

‘Onlogische keuzes’ met Artificial intelligence?

Met Artificial intelligence (AI) worden soms keuzes gemaakt die wij als onderzoekers betrokken bij dit project wellicht anders hadden gemaakt. Het AI-model kijkt enkel naar de getallen die zijn aangeleverd door de onderzoekers en maakt daartussen verbanden, terwijl bij een beoordeling van de onderzoekers de context een grote rol speelt. Daarnaast kijken de onderzoekers vanuit het perspectief voor klimaatadaptatie mogelijkheden en welke typering dan passend is. Een voorbeeld hiervan is het Singelpark in Utrecht, Figuur 15, waar ook deels gebouwen in staan. Ondanks een aanzienlijk deel groen is het geclassificeerd door het model als Historische Binnenstad, gebaseerd op de jaartallen van de gebouwen. Op PC6 niveau is de keuze tussen groen en historische binnenstad te bediscussiëren. Op PC5 en buurtniveau zijn deze PC6 gebiedjes zeker onderdeel van de historische binnenstad.



Figuur 15: PC6 gebieden met overwegend parkoppervlak die het AI-model labelt met het wijktipe Historische binnenstad.

7.2. Aanbevelingen voor verbeteringen

Voor de wijktypenkaart zijn waar mogelijk de meest geschikte datasets gebruikt en is gestuurd op een optimale verbetering in de voorspelling van wijktypen.

Bebouwd gebied

Door het bebouwd gebied in te delen in wijktypen kan inzicht worden verkregen in klimaatadaptatie mogelijkheden. Hierbij is het van groot belang het bebouwd gebied waar woningen staan goed te definiëren. De resultaten van het clusteren van PC6 gebieden naar PC5 gebieden en buurten laat zien dat grote infrastructuur en wateren een vertekend beeld kunnen geven over de toewijzing van een typologie. Dat wil zeggen dat als er veel wateren of grote wegen aanwezig zijn in een buurt, de typologie van het bebouwd gebied moeilijker te definiëren is. Voor een verdere verbetering van de kaart kan het zijn dat de grote infrastructuren en wateren uit de data moet worden verwijderd om nauwkeuriger de typologie van het bebouwd gebied te definiëren.

Bebouwde kom grenzen



In de verbetering van de kaartlaag die is gemaakt in 2017 zijn een aantal nieuwe datasets gemaakt. Een van deze datasets is de bebouwde kom. Deze laag geeft de ruimtelijke grens weer van de PC6 gebieden in Nederland die worden meegenomen in de classificatie. Hiervoor is samen met Stichting CAS een methodiek ontwikkeld, zoals beschreven in hoofdstuk 2. Destijds is erkend dat niet alle PC6 gebieden waar wel bebouwing in voorkomt mee zijn genomen. De afweging die destijds is genomen is dat bij een ruimere criteria ook veel groengebieden mee werden genomen. Uit een vergelijking met de meegenomen gebieden in de 2017 kaart is alsnog een aanzienlijke verbetering te constateren in de vernieuwde kaart.

Verharding

Dit project heeft gebruik gemaakt van BGT-gegevens om verharde oppervlakken te bepalen. Helaas is deze dataset beperkt tot alleen openbare verharde oppervlakken. Bovendien wordt de BGT handmatig gedigitaliseerd door gemeenten, wat betekent dat oppervlakken verkeerd geclassificeerd kunnen zijn. Om dit te verbeteren, zou het beter zijn om LiDAR-gegevens of luchtfoto's te gebruiken om verharde oppervlakken te bepalen.

Uitbreiding validatieset

Om het model nog beter te trainen zou een uitbreiding van de validatieset een verbetering kunnen opleveren. Hierbij is het van belang dat van de verschillende wijktypen een even groot aantal aanwezig zijn in de validatieset. Daarnaast moet ook de variatie van kleinere en grotere gemeenten gewaarborgd blijven. Zo leert het model de kenmerken van wijktypen in variërende context en kan het een betere voorspelling doen op nationaal niveau.

Weergave groen

Groen in de wijktypen speelt een belangrijke rol in de bepaling van het wijctype en vormt onderdeel van het type, maar wordt dan niet getoond als groen. Groen in de vorm van parken is soms onderdeel van een buurt, een wijk of structurerend op schaal van de stad. De aanbeveling is om deze als laag (bijv. TOP10NL) over de PC6 gebieden te leggen om zo de kaart beter herkenbaar te maken.

Literatuurlijst

Kleerekoper, L. (2016) Urban climate design: Improving thermal comfort in Dutch neighbourhoods. Delft University of Technology.

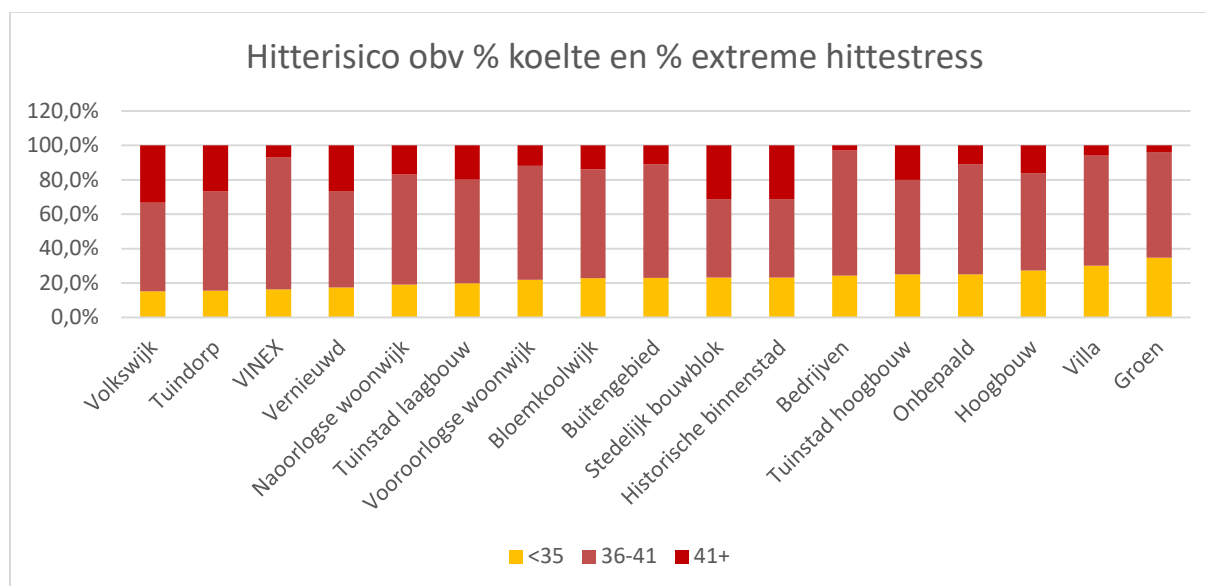
Kluck, J., Loeve, R., Bakker, W., Kleerekoper, L., Rouvoet, M., Wentink, R., Viscaal, J., Klok, L. and Boogaard, F. (2017) Het klimaat past ook in uw straatje: de waarde van klimaatbestendig inrichten. Voorbeeldenboek.

Kluck, J., Kleerekoper, L., Solcerová, A., Loeve, R., Erwin, S., Liu, C., Welter, N., Lopes, M., Rajaei, S. (2020) De hittebestendige stad: COOLKIT. Toolkit voor ontwerpers van de buitenruimte. Hogeschool van Amsterdam en KuiperCompagnons.

Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 5 June 2016.

Bijlage 1

De score voor 'Kwetsbaarheid hitte' per wijktype is gebaseerd op een analyse (door Deltares en de HvA) van de spreiding van de gevoelstemperatuur per wijktype, zie Figuur 12. Het percentage van het oppervlak met koelte (<35 graden PET) min het oppervlak met extreme hittestress (>41 graden PET) en bepaald de score voor de Kwetsbaarheid hitte in Tabel 6.



Figuur 16: spreiding van de gevoelstemperatuur per wijktype in een categorie koelte (<35 graden PET), matige hittestress (36-41 graden PET) en extreme hittestress (>41 graden PET)

Tabel 5: Kwetsbaarheid hitte

wijktype	<35	36-41	41+	koelte - extreme hitte	Kwetsbaarheid hitte
Volkswijk	0,15	0,52	0,33	-0,18	3
Tuindorp	0,15	0,58	0,27	-0,11	3
Vernieuwd	0,17	0,56	0,27	-0,09	3
Stedelijk bouwblok	0,23	0,45	0,31	-0,08	3
Historische binnenstad	0,23	0,45	0,31	-0,08	3
Tuinstad laagbouw	0,20	0,60	0,20	0,00	2
Naoorlogse woonwijk	0,19	0,64	0,17	0,02	2
Tuinstad hoogbouw	0,25	0,55	0,20	0,05	2
Bloemkoolwijk	0,23	0,63	0,14	0,09	2
VINEX	0,16	0,77	0,07	0,09	2
Vooroorlogse woonwijk	0,22	0,66	0,12	0,10	1
Hoogbouw	0,27	0,57	0,16	0,11	1
Buitengebied	0,23	0,66	0,11	0,12	1
Bedrijven	0,24	0,73	0,03	0,21	1
Villa	0,30	0,64	0,06	0,24	1
Groen	0,35	0,61	0,04	0,31	1

Bijlage 2

De wijktypenkaart bestaat uit drie verschillende datasets, namelijk de wijktypen op verschillende schaalniveaus: PC6, PC5 en Buurten. De tabel hieronder geeft een overzicht van de metadata, waaronder het datatype, tags, samenvatting, attribuutbeschrijving, 'credits' (makers van de datasets, gebruiksbeperking, omvang en schaalbereik).

Type	Shapefile
Tags	Wijktypen
Data Samenvatting	De inrichting van een Nederlandse straat is vaak kenmerkend voor een bepaalde tijd. De typerende eigenschappen van straten vinden we overal in Nederland terug. We kunnen daarom straten indelen in een typologie. Voor de Klimaat-effectatlas is een onderverdeling gemaakt die aansluit bij het voorbeeldenboek over klimaatbestendig inrichten van woonstraten van de Hogeschool van Amsterdam en dat is gebaseerd op de classificering van Kleerekoper (2016).
Beschrijving attribuut	Postcode: Ruimtelijk gebied voor elke postcode 6 uit de CSB PC6 2027 dataset. (Aanwezig in de PC6 attributentabel)
	Prediction: AI voorspelde het wijktipe op basis van verschillende inputs. (Aanwezig in de PC6 attributentabel)
	Probability: Een indicator voor hoe waarschijnlijk de wijktipe-voorspelling is. (Aanwezig in de PC6 attributentabel)
	Pred_aange: Aangepast wijktipe op basis van waarschijnlijkheid; als de waarschijnlijkheid kleiner was dan 0,2 dan werd de PC6 onbepaald. (Aanwezig in de PC6 attributentabel)
	PC5: Ruimtelijk gebied voor elke postcode 5 uit de CSB PC5 2017 dataset. De originele PC5-dataset bestond uit meerdelige poligonen. Deze polygonen zijn opgesplitst in afzonderlijke delen, maar behouden de oorspronkelijke unieke PC5-identificatie. (Aanwezig in de PC5 attributentabel)
	PC5_WT_ID: Unieke identificatie voor de PC5 wijktypen-dataset (Aanwezig in de PC6 attributentabel)
	wijktipe: Voorspelde wijktipe met de hoogste vertrouwenswaarde op basis van gewogen waarden. Voor meer informatie kunt u het projectrapport lezen. (Aanwezig in de attributentabel van PC5 en buurten)
	conf: Vertrouwen in het voorspelde wijktipe. Voor meer informatie kunt u het projectrapport lezen. (Aanwezig in de attributentabel van PC5 en buurten)
	wijktipe2: Voorspelde wijktipe met de op een na hoogste vertrouwenswaarde op basis van gewogen waarden. Voor meer informatie kunt u het



	projectrapport lezen. (Aanwezig in de attributentabel van PC5 en buurten)
	conf2: Vertrouwen in het voorspelde wijktype2. Voor meer informatie kunt u het projectrapport lezen. (Aanwezig in de attributentabel van PC5 en buurten)
	BU_CODE: Unieke identificatie van de CBS Wijk- en Buurtkaart 2022. Voor de codering van de binnen wijken onderscheiden buurten is een code van acht posities opgenomen. (Aanwezig in de buurten attributentabel)
	def-wijkty: Definitief wijktype na het forceren van de voorspelling met de volgende overweging: Als "wijktype" is "groen" en "conf" kleiner is dan of gelijk is aan 0,71 en "conf2" groter is dan of gelijk is aan 0,2 dan wordt "wijktype2" geforceerd in "def_wijktype"; de rest blijft behouden als het oorspronkelijke wijktype. (Aanwezig in de buurten attributentabel)
Credits	Hogeschool van Amsterdam en Stichting Climate Adaptation Services
Gebruiksbeperkingen	Er zijn geen toegangs- en gebruiksbeperkingen voor dit item.
Extent	Er is geen extent voor dit item.
Schaal Bereik	Er is geen schaalbereik voor dit item.